Supplementary material for
Continuous changepoint monitoring of data streams using adaptive estimation

Dean A. Bodenham Niall M. Adams

Statistics and Computing

1 Proofs of propositions regarding the computation of $\bar{x}_{N,\overrightarrow{\lambda}}$

This section contains the proofs for several propositions which were stated in the main text. For convenience, they are stated again here before being proved, and the initial definitions are reproduced here:

Definition 1. *For a sequence of observations $x_1, x_2, \ldots, x_N$, and forgetting factors $\lambda_1, \lambda_2, \ldots, \lambda_N$, after defining*

$$m_{N+1,\overrightarrow{\lambda}} = \lambda_N m_{N,\overrightarrow{\lambda}} + x_{N+1}, \qquad m_{0,\overrightarrow{\lambda}} = 0, \tag{1}$$
$$w_{N+1,\overrightarrow{\lambda}} = \lambda_N w_{N,\overrightarrow{\lambda}} + 1, \qquad w_{0,\overrightarrow{\lambda}} = 0,$$

the adaptive forgetting factor mean $\bar{x}_{N,\overrightarrow{\lambda}}$ is defined, for $N \geq 1$, by $\bar{x}_{N,\overrightarrow{\lambda}} = \frac{m_{N,\overrightarrow{\lambda}}}{w_{N,\overrightarrow{\lambda}}}$.

Definition 2. *For any function $f_{N,\overrightarrow{\lambda}}$ involving $\overrightarrow{\lambda}$,*

$$\frac{\partial}{\partial \overrightarrow{\lambda}} f_{N,\overrightarrow{\lambda}} = \lim_{\epsilon \to 0} \left[f_{N,\overrightarrow{\lambda}+\epsilon} - f_{N,\overrightarrow{\lambda}} \right],$$

where $\overrightarrow{\lambda} + \epsilon = (\lambda_1 + \epsilon, \lambda_2 + \epsilon, \ldots)$.

1.1 Proposition 3: Non-sequential form for $m_{N,\overrightarrow{\lambda}}$

The following proposition is needed to define the derivative of $m_{N,\overrightarrow{\lambda}}$ with respect to $\overrightarrow{\lambda}$:

Proposition 3. *For $N = 1, 2, \ldots$, $m_{N,\overrightarrow{\lambda}}$ defined sequentially in Definition 1 can be computed using:*

$$m_{N,\overrightarrow{\lambda}} = \sum_{i=1}^{N} \left[\left(\prod_{p=i}^{N-1} \lambda_p \right) x_i \right]. \tag{2}$$

Proof:

First note that the sequential definition of $m_{N,\overrightarrow{\lambda}}$ in Equation (1) can be rewritten as:

$$m_{N,\overrightarrow{\lambda}} = \lambda_{N-1} m_{N-1,\overrightarrow{\lambda}} + x_N, \qquad N \geq 1, \qquad m_{0,\overrightarrow{\lambda}} = 0, \tag{3}$$

Now, the easiest way to prove this proposition is to observe that the function

$$f(N) \equiv f(N, \lambda_1, \ldots, \lambda_{N-1}, x_1, \ldots, x_N) = \sum_{i=1}^{N} \left[\Big(\prod_{p=i}^{N-1} \lambda_p \Big) x_N \right], \tag{4}$$

satisfies the relation

$$f(N) = \lambda_{N-1} f(N-1) + x_N. \tag{5}$$

which mirrors the sequential definition of $m_{N,\overrightarrow{\lambda}}$ in Equation (3). Recalling the fact that the empty product is 1, i.e.

$$\prod_{N=M}^{M-1} (x_N) = 1,$$

we have:

$$\begin{aligned}
f(N) &= \sum_{i=1}^{N} \Big[\Big(\prod_{p=i}^{N-1} \lambda_p \Big) x_N \Big] \\
&= \sum_{i=1}^{N-1} \Big[\Big(\prod_{p=i}^{N-1} \lambda_p \Big) x_N \Big] + \sum_{i=N}^{N} \Big[\Big(\prod_{p=i}^{N-1} \lambda_p \Big) x_N \Big] \\
&= \sum_{i=1}^{N-1} \Big[\Big(\prod_{p=i}^{N-1} \lambda_p \Big) x_N \Big] + \Big(\prod_{p=N}^{N-1} \lambda_p \Big) x_N \\
&= \sum_{i=1}^{N-1} \Big[\Big(\prod_{p=i}^{N-1} \lambda_p \Big) x_N \Big] + x_N \\
&= \sum_{i=1}^{N-1} \Big[\lambda_{N-1} \Big(\prod_{p=i}^{N-2} \lambda_p \Big) x_N \Big] + x_N \\
&= \lambda_{N-1} \sum_{i=1}^{N-1} \Big[\Big(\prod_{p=i}^{N-2} \lambda_p \Big) x_N \Big] + x_N \\
\Rightarrow f(N) &= \lambda_{N-1} f(N-1) + x_N
\end{aligned} \tag{6}$$

which proves that Equation (4) satisfies the relation in Equation (5). This proves that $m_{N,\overrightarrow{\lambda}}$ can be computed using Equation (2), which proves the proposition. □

1.1.1 Remark

Note that by following a similar calculation one can easily show

$$w_{N,\overrightarrow{\lambda}} = \sum_{i=1}^{N} \Big(\prod_{p=i}^{N-1} \lambda_p \Big).$$

1.2 Lemma 4: A relation needed for the computation of the derivative with respect to $\overrightarrow{\lambda}$

The following lemma is needed in order to calculate the derivative of quantities involving $m_{k,\overrightarrow{\lambda}}$ and $w_{k,\overrightarrow{\lambda}}$:

Lemma 4. *For $\lambda_i, \lambda_{i+1}, \ldots, \lambda_M$, $i \geq 1$, and $\epsilon \ll 1$,*

$$\prod_{t=i}^{M} \left(\lambda_t + \epsilon\right) = \prod_{t=i}^{M} \lambda_t + \epsilon \left(\sum_{t=i}^{M} \left(\prod_{\substack{p=i \\ p \neq t}}^{M} \lambda_p \right) \right) + O(\epsilon^2). \tag{7}$$

In fact, this lemma is a special case of a more general result, which we name Lemma 4*. We first state and prove Lemma 4* before proving Lemma 4.

Lemma 4*. *For sequences $a_i, a_{i+1}, \ldots, a_M$ and $b_i, b_{i+1}, \ldots, b_M$, and $\epsilon \ll 1$,*

$$\prod_{t=i}^{M} (a_t + b_t\epsilon) = \prod_{t=i}^{M} a_t + \epsilon \sum_{t=i}^{M} \left(b_t \prod_{\substack{p=i \\ p \neq t}}^{M} a_p \right) + O(\epsilon^2)$$

Proof (Lemma 4*):

We prove the result by induction. Note that M is simply an upper limit greater than or equal to i.

Case 1: $M = i$

$$\begin{aligned}
\text{LHS} &= \prod_{t=i}^{i} (a_t + b_t\epsilon) \\
&= a_i + b_i\epsilon \\
&= a_i + b_i\epsilon + O(\epsilon^2) \\
\text{RHS} &= \prod_{t=i}^{i} a_t + \epsilon \sum_{t=i}^{i} \left(b_t \prod_{\substack{p=i \\ p\neq t}}^{i} a_p \right) + O(\epsilon^2) \\
&= a_i + \epsilon \left(b_i \prod_{\substack{p=i \\ p\neq i}}^{i} a_p \right) + O(\epsilon^2) \\
&= a_i + \epsilon b_i(1) + O(\epsilon^2) \\
&= a_i + b_i\epsilon + O(\epsilon^2) \\
&= \text{LHS}
\end{aligned}$$

using the fact that the empty product is 1.

Assumption: We first assume that the result holds for $M = N$,

$$\prod_{t=i}^{N} (a_t + b_t\epsilon) = \prod_{t=i}^{N} a_t + \epsilon \sum_{t=i}^{N} \left(b_t \prod_{\substack{p=i \\ p\neq t}}^{N} a_p \right) + O(\epsilon^2)$$

Case 2: $M = N + 1$

$$\begin{aligned}
\text{RHS} &= \prod_{t=i}^{N+1} a_t + \epsilon \sum_{t=i}^{N+1} \left(b_t \prod_{\substack{p=i \\ p\neq t}}^{N+1} a_p \right) + O(\epsilon^2) \\
\text{LHS} &= \prod_{t=i}^{N+1} (a_t + b_t\epsilon) \\
&= \left[\prod_{t=i}^{N} (a_t + b_t\epsilon) \right] (a_{N+1} + b_{N+1}\epsilon) \\
&= \left[\prod_{t=i}^{N} a_t + \epsilon \sum_{t=i}^{N} \left(b_t \prod_{\substack{p=i \\ p\neq t}}^{N} a_p \right) + O(\epsilon^2) \right] (a_{N+1} + b_{N+1}\epsilon) \\
&= \left(\prod_{t=i}^{N} a_t \right) (a_{N+1} + b_{N+1}\epsilon) + \epsilon \sum_{t=i}^{N} \left(b_t \prod_{\substack{p=i \\ p\neq t}}^{N} a_p \right) (a_{N+1} + b_{N+1}\epsilon) + O(\epsilon^2) \\
&= \left(\prod_{t=i}^{N} a_t \right) (a_{N+1} + b_{N+1}\epsilon) + \epsilon \sum_{t=i}^{N} \left(b_t \prod_{\substack{p=i \\ p\neq t}}^{N} a_p \right) (a_{N+1}) + O(\epsilon^2) \\
&= \left(\prod_{t=i}^{N+1} a_t \right) + \epsilon b_{N+1} \left(\prod_{t=i}^{N} a_t \right) + \epsilon \sum_{t=i}^{N} \left(b_t a_{N+1} \prod_{\substack{p=i \\ p\neq t}}^{N} a_p \right) + O(\epsilon^2) \\
&= \left(\prod_{t=i}^{N+1} a_t \right) + \epsilon b_{N+1} \left(\prod_{\substack{p=i \\ p\neq N+1}}^{N+1} a_p \right) + \epsilon \sum_{t=i}^{N} \left(b_t \prod_{\substack{p=i \\ p\neq t}}^{N+1} a_p \right) + O(\epsilon^2) \\
&= \left(\prod_{t=i}^{N+1} a_t \right) + \epsilon \sum_{t=N+1}^{N+1} \left(b_t \prod_{\substack{p=i \\ p\neq t}}^{N+1} a_p \right) + \epsilon \sum_{t=i}^{N} \left(b_t \prod_{\substack{p=i \\ p\neq t}}^{N+1} a_p \right) + O(\epsilon^2) \\
&= \left(\prod_{t=i}^{N+1} a_t \right) + \epsilon \sum_{t=i}^{N+1} \left(b_t \prod_{\substack{p=i \\ p\neq t}}^{N+1} a_p \right) + O(\epsilon^2) \\
&= \text{RHS}
\end{aligned}$$

Therefore, by induction the result holds for any $M \geq i$. This completes the proof of Lemma 4*. □

Remark:

Although a direct proof is possible, the proof by induction is perhaps a bit better.

Proof (of Lemma 4):

Using Lemma 4*, with

$$\begin{aligned} a_t &= \lambda_i \\ b_t &= 1 \end{aligned}$$

for $t = i, i+1, \ldots M$, we immediately have:

$$\prod_{t=i}^{M} (\lambda_i + \epsilon) = \prod_{t=i}^{M} \lambda_t + \epsilon \sum_{t=i}^{M} \left(\prod_{\substack{p=i \\ p\neq t}}^{M} \lambda_p \right) + O(\epsilon^2).$$

Although it is trivial, relabelling the iterators in the product on the left-hand side, and the first product on the right-hand side, from t to p, we have it in exactly the same form as in Lemma 4:

$$\prod_{p=i}^{M}(\lambda_i+\epsilon)=\prod_{p=i}^{M}\lambda_p+\epsilon\sum_{t=i}^{M}\left(\prod_{\substack{p=i\\p\neq t}}^{M}\lambda_p\right)+O(\epsilon^2).$$

This completes the proof of Lemma 4. $\square$

1.2.1 Remark

A very convenient form of the lemma is

$$\prod_{t=i}^{M}\left(\lambda_t+\epsilon\right)=\prod_{t=i}^{M}\lambda_t+\epsilon\left(\sum_{t=i}^{M}\left(\prod_{\substack{p=i\\p\neq t}}^{M}\lambda_p\right)\right)+O(\epsilon^2). \tag{8}$$

1.3 The definition of $\Delta_{k,\overrightarrow{\lambda}}$, derivative of $m_{k,\overrightarrow{\lambda}}$ with respect to $\overrightarrow{\lambda}$

Recall the definition of the derivative of $m_{k,\overrightarrow{\lambda}}$ with respect to $\overrightarrow{\lambda}$:

$$\Delta_{N,\overrightarrow{\lambda}} = \frac{\partial}{\partial \overrightarrow{\lambda}} m_{N,\overrightarrow{\lambda}} = \lim_{\epsilon \to 0} \frac{1}{\epsilon} \Big[m_{N,\overrightarrow{\lambda}+\epsilon} - m_{N,\overrightarrow{\lambda}} \Big]. \tag{9}$$

This leads to the following proposition:

Proposition 5. *Following the definition of $\Delta_{N,\overrightarrow{\lambda}}$ in Equation* (9)*, $\Delta_{N,\overrightarrow{\lambda}}$ can be computed using:*

$$\Delta_{N,\overrightarrow{\lambda}} = \sum_{i=1}^{N-1} \Big[\sum_{t=i}^{N-1} \Big(\prod_{\substack{p=i \\ p\neq t}}^{N-1} \lambda_p \Big) x_i \Big]. \tag{10}$$

Proof:

Using the definition of the derivative with respect to $\overrightarrow{\lambda}$, and the non-sequential form of $m_{k,\overrightarrow{\lambda}}$ proved in Proposition 3,

$$\begin{aligned}
\Delta_{N,\overrightarrow{\lambda}} &= \lim_{\epsilon \to 0} \frac{1}{\epsilon} \Big[m_{N,\overrightarrow{\lambda}+\epsilon} - m_{N,\overrightarrow{\lambda}} \Big] \\
&= \lim_{\epsilon \to 0} \frac{1}{\epsilon} \left[\sum_{i=1}^{N} \left[\left(\prod_{p=i}^{N-1} (\lambda_p + \epsilon) \right) x_i \right] - \sum_{i=1}^{N} \left[\left(\prod_{p=i}^{N-1} \lambda_p \right) x_i \right] \right] \\
&= \lim_{\epsilon \to 0} \frac{1}{\epsilon} \sum_{i=1}^{N} \left[\prod_{p=i}^{N-1} (\lambda_p + \epsilon) - \prod_{p=i}^{N-1} \lambda_p \right] x_i
\end{aligned}$$

Now, using Lemma 4,

$$\begin{aligned}
\Delta_{N,\overrightarrow{\lambda}} &= \lim_{\epsilon \to 0} \frac{1}{\epsilon} \sum_{i=1}^{N} \left[\epsilon \left(\sum_{j=i}^{N-1} \Big(\prod_{\substack{p=i \\ p\neq j}}^{N-1} \lambda_p \Big) \right) + O(\epsilon^2) \right] x_i \qquad (11) \\
&= \lim_{\epsilon \to 0} \sum_{i=1}^{N} \left[\frac{1}{\epsilon} \cdot \epsilon \left(\sum_{j=i}^{N-1} \Big(\prod_{\substack{p=i \\ p\neq j}}^{N-1} \lambda_p \Big) \right) + \frac{1}{\epsilon} \cdot O(\epsilon^2) \right] x_i \\
&= \lim_{\epsilon \to 0} \sum_{i=1}^{N} \left[\sum_{j=i}^{N-1} \Big(\prod_{\substack{p=i \\ p\neq j}}^{N-1} \lambda_p \Big) + O(\epsilon) \right] x_i \\
\Rightarrow \Delta_{N,\overrightarrow{\lambda}} &= \sum_{i=1}^{N} \left[\sum_{j=i}^{N-1} \Big(\prod_{\substack{p=i \\ p\neq j}}^{N-1} \lambda_p \Big) x_i \right]. \qquad (12)
\end{aligned}$$

Finally, we show Equation (12) is equivalent to

$$\Delta_{N,\overrightarrow{\lambda}} = \sum_{i=1}^{N-1} \Big[\sum_{j=i}^{N-1} \Big(\prod_{\substack{p=i \\ p\neq j}}^{N-1} \lambda_p \Big) x_i \Big], \tag{13}$$

(where the first summation over i runs over $i = 1, \ldots, N$ or $i = 1, \ldots, N-1$), since

$$\begin{aligned}
\sum_{i=1}^{N}\left[\sum_{j=i}^{N-1}\Big(\prod_{\substack{p=i\\p\neq j}}^{N-1}\lambda_p\Big)x_i\right] &= \sum_{i=1}^{N-1}\left[\sum_{j=i}^{N-1}\Big(\prod_{\substack{p=i\\p\neq j}}^{N-1}\lambda_p\Big)x_i\right] + \sum_{i=N}^{N}\left[\sum_{j=i}^{N-1}\Big(\prod_{\substack{p=i\\p\neq j}}^{N-1}\lambda_p\Big)x_i\right]\\
&= \sum_{i=1}^{N-1}\left[\sum_{j=i}^{N-1}\Big(\prod_{\substack{p=i\\p\neq j}}^{N-1}\lambda_p\Big)x_i\right] + \left[\sum_{j=N}^{N-1}\Big(\prod_{\substack{p=i\\p\neq j}}^{N-1}\lambda_p\Big)x_i\right]\\
&= \sum_{i=1}^{N-1}\left[\sum_{j=i}^{N-1}\Big(\prod_{\substack{p=i\\p\neq j}}^{N-1}\lambda_p\Big)x_i\right] + 0
\end{aligned}$$

where we have used the fact that an empty summation

$$\sum_{j=N}^{N-1}\alpha_j = 0 \tag{14}$$

(for example, a sum from a higher to a lower index) must be zero. □

1.3.1 Remark

It can be similarly shown that

$$\Omega_{N,\overrightarrow{\lambda}} = \frac{\partial}{\partial\overrightarrow{\lambda}} w_{N,\overrightarrow{\lambda}} = \sum_{i=1}^{N-1}\left[\sum_{t=i}^{N-1}\Big(\prod_{\substack{p=i\\p\neq t}}^{N-1}\lambda_p\Big)\right]. \tag{15}$$

1.4 Sequential update equations for $\Delta_{k,\overrightarrow{\lambda}}$ and $\Omega_{k,\overrightarrow{\lambda}}$

We now derive a sequential update equation of $\Delta_{N,\overrightarrow{\lambda}}$, which is stated in the following proposition:

Proposition. *Using the computation of $\Delta_{N,\overrightarrow{\lambda}}$ in Proposition 5, $\Delta_{N,\overrightarrow{\lambda}}$ can be computed sequentially using:*

$$\Delta_{N+1,\overrightarrow{\lambda}} = \lambda_N\Delta_{N,\overrightarrow{\lambda}} + m_{N,\lambda}, \qquad \Delta_{1,\overrightarrow{\lambda}} = 0 \tag{16}$$

Proof:

We compute the left- and right-hand sides of Equation (16) using the non-sequential form of $\Delta_{N,\overrightarrow{\lambda}}$ in Equation (10) and show the expressions are equivalent:

$$\begin{aligned}
\text{LHS} = \Delta_{N+1,\overrightarrow{\lambda}} &= \sum_{k=1}^{N+1}\left[\sum_{t=k}^{N}\Big(\prod_{\substack{p=k\\p\neq t}}^{N}\lambda_p\Big)x_k\right]\\
&= \sum_{k=1}^{N}\left[\sum_{t=k}^{N}\Big(\prod_{\substack{p=k\\p\neq t}}^{N}\lambda_p\Big)x_k\right]
\end{aligned}$$

$$
\begin{aligned}
\text{RHS} &= \lambda_N \Delta_{N,\overrightarrow{\lambda}} + m_{N,\lambda} \\
&= \lambda_N \sum_{k=1}^{N} \left[\sum_{t=k}^{N-1} \Big(\prod_{\substack{p=k \\ p\neq t}}^{N-1} \lambda_p \Big) x_k \right] + \sum_{k=1}^{N} \left[\Big(\prod_{p=k}^{N-1} \lambda_p \Big) x_k \right] \\
&= \sum_{k=1}^{N} \left[\sum_{t=k}^{N-1} \lambda_N \Big(\prod_{\substack{p=k \\ p\neq t}}^{N-1} \lambda_p \Big) x_k \right] + \sum_{k=1}^{N} \left[\Big(\prod_{p=k}^{N-1} \lambda_p \Big) x_k \right] \\
&= \sum_{k=1}^{N} \left[\sum_{t=k}^{N-1} \Big(\prod_{\substack{p=k \\ p\neq t}}^{N} \lambda_p \Big) x_k \right] + \sum_{k=1}^{N} \left[\Big(\prod_{p=k}^{N-1} \lambda_p \Big) x_k \right] \\
&= \sum_{k=1}^{N} \left[\sum_{t=k}^{N-1} \Big(\prod_{\substack{p=k \\ p\neq t}}^{N} \lambda_p \Big) + \Big(\prod_{p=k}^{N-1} \lambda_p \Big) \right] x_k \\
&= \sum_{k=1}^{N} \left[\sum_{t=k}^{N} \Big(\prod_{\substack{p=k \\ p\neq t}}^{N} \lambda_p \Big) - \sum_{t=N}^{N} \Big(\prod_{\substack{p=k \\ p\neq t}}^{N} \lambda_p \Big) + \Big(\prod_{p=k}^{N-1} \lambda_p \Big) \right] x_k \\
&= \sum_{k=1}^{N} \left[\sum_{t=k}^{N} \Big(\prod_{\substack{p=k \\ p\neq t}}^{N} \lambda_p \Big) - \Big(\prod_{\substack{p=k \\ p\neq N}}^{N} \lambda_p \Big) + \Big(\prod_{p=k}^{N-1} \lambda_p \Big) \right] x_k \\
&= \sum_{k=1}^{N} \left[\sum_{t=k}^{N} \Big(\prod_{\substack{p=k \\ p\neq t}}^{N} \lambda_p \Big) - \Big(\prod_{p=k}^{N-1} \lambda_p \Big) + \Big(\prod_{p=k}^{N-1} \lambda_p \Big) \right] x_k \\
&= \sum_{k=1}^{N} \left[\sum_{t=k}^{N} \Big(\prod_{\substack{p=k \\ p\neq t}}^{N} \lambda_p \Big) \right] x_k \\
&= \text{LHS}
\end{aligned}
$$

which proves the relation. □

1.4.1 Remark

Following a similar proof, it can be shown that $\Omega_{N+1,\overrightarrow{\lambda}} = \lambda_N \Omega_{N,\overrightarrow{\lambda}} + w_{N,\lambda}$.

1.5 Derivative of cost function calculation

In Section 4.2.2, the derivative of the cost function $L_{k,\overrightarrow{\lambda}} = [\bar{x}_{k-1,\overrightarrow{\lambda}} - x_k]^2$ is given. Here is the calculation:

$$\begin{aligned}
\frac{\partial}{\partial \overrightarrow{\lambda}} L_{k,\overrightarrow{\lambda}} &= \frac{\partial}{\partial \overrightarrow{\lambda}} [\bar{x}_{k-1,\overrightarrow{\lambda}} - x_k]^2 \\
&= 2[\bar{x}_{k-1,\overrightarrow{\lambda}} - x_k] \left(\frac{\partial}{\partial \overrightarrow{\lambda}} [\bar{x}_{k-1,\overrightarrow{\lambda}} - x_k] \right) \\
&= 2[\bar{x}_{k-1,\overrightarrow{\lambda}} - x_k] \left(\frac{\partial}{\partial \overrightarrow{\lambda}} \left[\frac{m_{k-1,\overrightarrow{\lambda}}}{w_{k-1,\overrightarrow{\lambda}}} \right] \right) \\
&= 2[\bar{x}_{k-1,\overrightarrow{\lambda}} - x_k] \left(\frac{\Delta_{k-1,\overrightarrow{\lambda}} w_{k-1,\overrightarrow{\lambda}} - m_{k-1,\overrightarrow{\lambda}} \Omega_{k-1,\overrightarrow{\lambda}}}{(w_{k-1,\overrightarrow{\lambda}})^2} \right) \\
&= 2[\bar{x}_{k-1,\overrightarrow{\lambda}} - x_k] \left(\frac{\Delta_{k-1,\overrightarrow{\lambda}} - \frac{m_{k-1,\overrightarrow{\lambda}}}{w_{k-1,\overrightarrow{\lambda}}} \Omega_{k-1,\overrightarrow{\lambda}}}{w_{k-1,\overrightarrow{\lambda}}} \right) \\
\Rightarrow \frac{\partial}{\partial \overrightarrow{\lambda}} L_{k,\overrightarrow{\lambda}} &= 2[\bar{x}_{k-1,\overrightarrow{\lambda}} - x_k] \left[\frac{\Delta_{k-1,\overrightarrow{\lambda}} - \bar{x}_{k-1,\overrightarrow{\lambda}} \Omega_{k-1,\overrightarrow{\lambda}}}{w_{k-1,\overrightarrow{\lambda}}} \right].
\end{aligned}$$

1.6 The effect of the value of $\lambda_{\mathbf{min}}$ on $\overrightarrow{\lambda}$

In Section 4.2.2, it is commented that we need to constrain each updated λ_k value to be in the interval $[0, 1]$ using

$$\lambda_k = \max\{\min\{\lambda_k, 1\}, \lambda_{min}\}, \qquad \lambda_{min} \in [0, 1)$$

because the gradient descent updating procedure

$$\lambda_k = \lambda_{k-1} - \eta \frac{\partial}{\partial \overrightarrow{\lambda}} L_{k,\overrightarrow{\lambda}},$$

may result in a value $\lambda_k \notin [0, 1]$. It was further commented that a value $\lambda_{min} = 0.6$ was recommended, and this section discusses the motivation for this choice.

1.6.1 The recovery of $\overrightarrow{\lambda}$ after a change

The primary motivation for this choice can be expressed by Figure 1, which shows the recovery of $\overrightarrow{\lambda}$ after a changepoint to pre-change levels. The left-hand panel shows that truncating $\overrightarrow{\lambda}$ at 0.6 allows for a relatively swift recovery, while the right-hand panel shows that allowing $\overrightarrow{\lambda}$ to decrease to 0.01 (without truncation) leads to a very slow recovery.

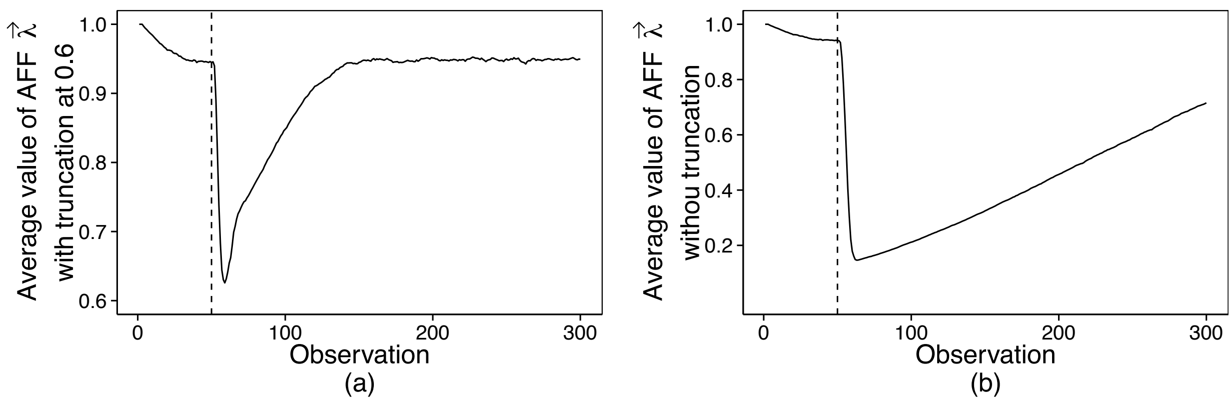

Figure 1: The average value of $\overrightarrow{\lambda}$ with truncation at (a) $\overrightarrow{\lambda}_{\text{min}} = 0.6$ and (b) $\overrightarrow{\lambda}_{\text{min}} = 0.01$. The average is obtained from 1000 streams $x_1, x_2, \ldots, x_{300}$ each sampled from $X_1, \ldots, X_{50} \sim \mathrm{N}(0, 1)$ and $X_{51}, \ldots, X_{300} \sim \mathrm{N}(2, 1)$. The step-size is $\eta = 0.01$.

However, Figure 1 is not enough, and one may consider the value $\lambda_{min} = 0.6$ to be a bit arbitrary (why not $\lambda_{min} = 0.5$ or $\lambda_{min} = 0.7$?). In order to understand this choice, one needs to consider the quantities $w_{k,\overrightarrow{\lambda}}$ and $u_{k,\overrightarrow{\lambda}}$.

1.6.2 The effective sample size and the variance of the $\bar{x}_{N,\overrightarrow{\lambda}}$

We recall that $w_{k,\overrightarrow{\lambda}}$ is part of the definition of the forgetting factor mean:

$$\bar{x}_{k,\overrightarrow{\lambda}} = \frac{m_{k,\overrightarrow{\lambda}}}{w_{k,\overrightarrow{\lambda}}} = \frac{1}{w_{k,\overrightarrow{\lambda}}} \sum_{i=1}^{k} \left[\left(\prod_{p=i}^{k-1} \lambda_p \right) x_i \right]$$

and when compared to the sample mean:

$$\bar{x}_k = \frac{1}{k} \sum_{i=1}^{k} x_i$$

one can (and should) consider the quantity $w_{k,\overrightarrow{\lambda}}$ to be the *effective sample size*, with $\overrightarrow{\lambda}$ controlling the size of this effective sample size. We can look at the case of a *fixed* forgetting factor λ, in order to gain some insight into this quantity. Recall from Section 4.1 that $w_{k,\lambda}$ was defined as:

$$w_{N,\lambda} = \sum_{i=1}^{N} \lambda^{N-i} = \frac{1-\lambda^N}{1-\lambda},$$

where the second equality is simply the formula for a geometric progression in λ. However, when $N \to \infty$, this results in:

$$w_{\infty,\lambda} = \lim_{N \to \infty} w_{N,\lambda} = \frac{1}{1-\lambda},$$

And so we can see how, in the limit, $\lambda = 0.95$ results in an effective sample size of 20.

In Section 4.3, Proposition 6 states that the variance of the theoretical forgetting factor mean is (see the next section in this Supplementary Material for the proof):

$$\mathrm{Var}[\bar{X}_{N,\overrightarrow{\lambda}}] = (u_{N,\overrightarrow{\lambda}})\sigma^2.$$

Therefore, the quantity $u_{N,\overrightarrow{\lambda}}$ controls the variance of $\bar{x}_{N,\overrightarrow{\lambda}}$, and a larger $u_{N,\overrightarrow{\lambda}}$ means a larger variance. A fixed forgetting factor version, $u_{N,\lambda}$ can simply be defined by fixing all values in $\overrightarrow{\lambda}$ to be a fixed λ.

Now consider, Figure 2 which shows that for small values of λ, $w_{N,\lambda}$ is very small and $u_{N,\lambda}$ is large, which means the variance of the estimator $\bar{x}_{N,\lambda}$ is large. Furthermore, Figure 2 shows that when λ is below 0.5, the values of $w_{N,\lambda}$ and $u_{N,\overrightarrow{\lambda}}$ for various N are virtually identical. However, for values above 0.75, both $w_{N,\lambda}$ and $u_{N,\overrightarrow{\lambda}}$ differ for increasing values of N. For this reason, Figure 3 reproduces Figure 2, but only showing the values of $w_{N,\lambda}$ and $u_{N,\overrightarrow{\lambda}}$ for $\lambda \in [0.6, 1.0]$, to allow the differences in this region to be better observed, and to show that 0.6 seems to be the point where both $w_{N,\lambda}$ and $u_{N,\lambda}$ seem to have a value that is unaffected by the choice of N.

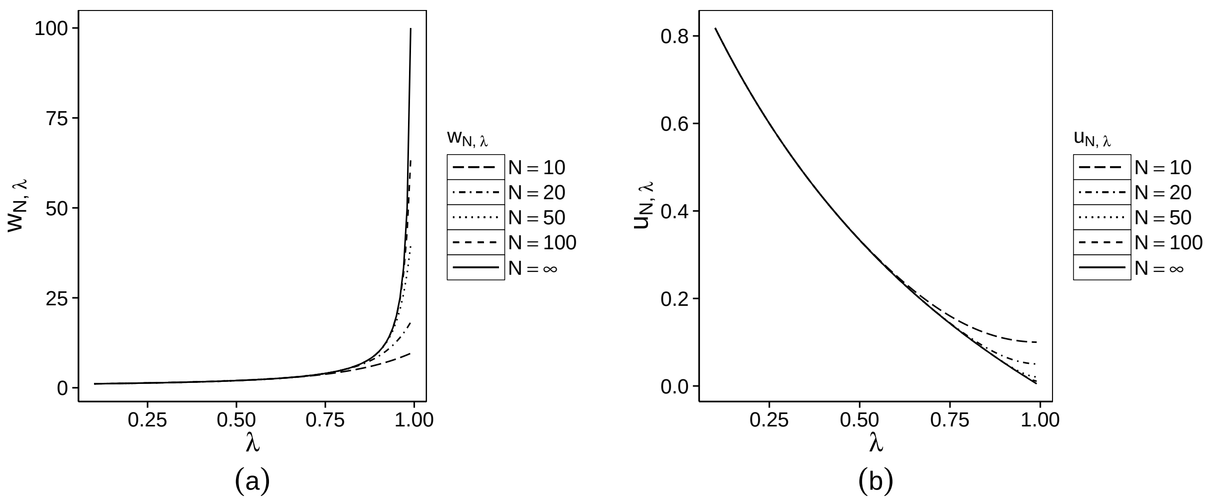

Figure 2: Values of (a) $w_{N,\lambda}$ and (b) $u_{N,\lambda}$ for various values of N, for $\lambda \in [0.1, 0.99]$.

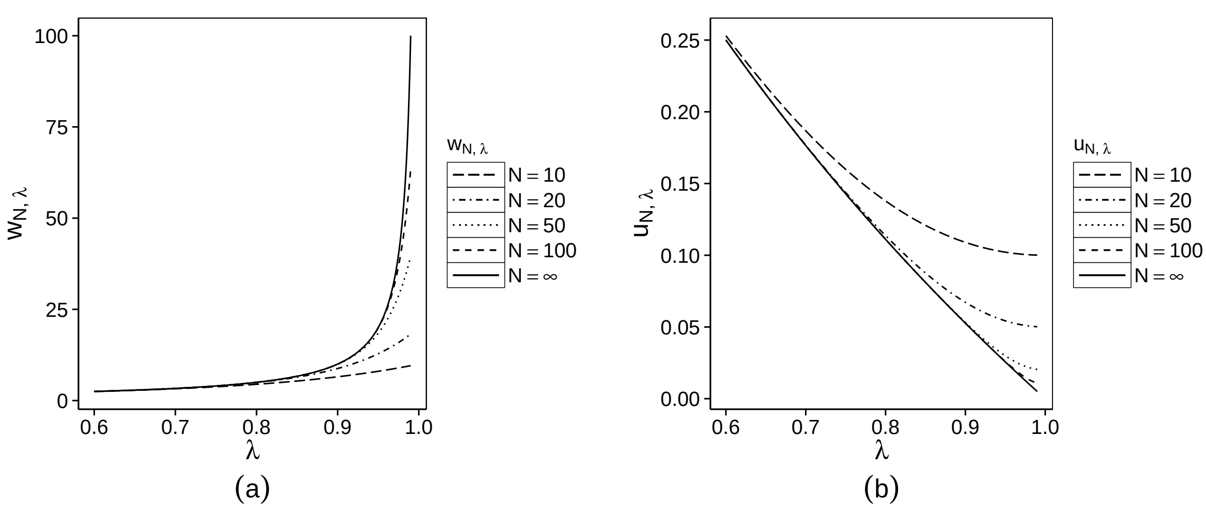

Figure 3: Values of (a) $w_{N,\lambda}$ and (b) $u_{N,\lambda}$for various values of N, for $\lambda \in [0.6, 0.99]$.

Furthermore, Figure 4 shows that truncating $\overrightarrow{\lambda}$ at $\overrightarrow{\lambda}_{\min} = 0.6$ keeps the value of $u_{N,\overrightarrow{\lambda}}$ low after a changepoint, in contrast to allowing $\overrightarrow{\lambda}$ to decrease to 0.01. The initial spike in $u_{N,\overrightarrow{\lambda}}$ is due to the estimation at the beginning of the burn-in having very few samples.

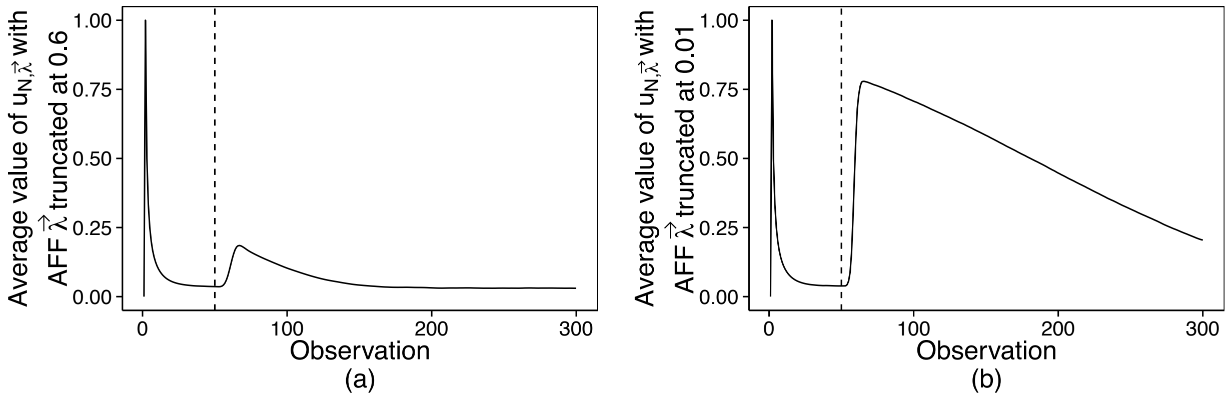

Figure 4: The average value of $u_{N,\overrightarrow{\lambda}}$ with truncation at (a) $\overrightarrow{\lambda}_{\min} = 0.6$ and (b) $\overrightarrow{\lambda}_{\min} = 0.01$. The average is obtained from 1000 streams $x_1, x_2, \ldots, x_{300}$ each sampled from $X_1, \ldots, X_{50} \sim \mathrm{N}(0, 1)$ and $X_{51}, \ldots, X_{300} \sim \mathrm{N}(2, 1)$. The step-size is $\eta = 0.01$.

This concludes this section, which provided some evidence that motivated out choice for truncating λ at $\lambda_{min} = 0.6$.

2 Expectation and variance of quantities involving $\bar{X}_{N,\overrightarrow{\lambda}}$

2.1 Derivation of the expectation and variance of $\bar{X}_{N,\overrightarrow{\lambda}}$

This section contains the proof of Proposition 6, which is used in order to derive a decision rule for detecting a change. Note that there is no assumption of normality, just that the X_i are i.i.d. with the same mean and variance.

Proposition 6. *If our data stream is sampled from the i.i.d. random variables $X_1, X_2, \ldots, X_N$, with $E[X_i] = \mu$ and $Var[X_i] = \sigma^2$ for all $i \geq 1$, then $\bar{X}_{N,\overrightarrow{\lambda}}$, the adaptive forgetting factor mean of $X_1, X_2, \ldots X_N$, defined according to Definition 1, has expectation and variance*

$$E[\bar{X}_{N,\overrightarrow{\lambda}}] = \mu, \qquad Var[\bar{X}_{N,\overrightarrow{\lambda}}] = (u_{N,\overrightarrow{\lambda}})\sigma^2,$$

where $u_{1,\overrightarrow{\lambda}} = 1$ and, for $i \geq 1$,

$$u_{i+1,\lambda} = \left(1 - \frac{1}{w_{i+1,\lambda}}\right)^2 u_{i,\lambda} + \left(\frac{1}{w_{i+1,\lambda}}\right)^2.$$

Proof:

This proof is split into two parts. In Part (a) the expectation and variance of $\bar{X}_{N,\lambda}$ are computed. In Part (b) the sequential update equation of $u_{N,\overrightarrow{\lambda}}$ is derived.

Part (a): Suppose that the random variables X_i are independent and identically-distributed (i.i.d.) with expectation and variance:

$$\mathrm{E}\left[X_i\right] = \mu, \qquad \mathrm{Var}\left[X_i\right] = \sigma^2.$$

Recall that $\bar{X}_{N,\lambda}$ is defined by:

$$\bar{X}_{N,\lambda} = \frac{m_{N,\overrightarrow{\lambda}}}{w_{N,\overrightarrow{\lambda}}}$$

where $m_{N,\overrightarrow{\lambda}}$ and $w_{N,\overrightarrow{\lambda}}$ are defined sequentially. Proposition 3 defined $m_{\mathrm{N},\overrightarrow{\lambda}}$, for observations $x_1, x_2, \ldots, x_k$:

$$m_{k,\overrightarrow{\lambda}} = \sum_{i=1}^{k} \left[\left(\prod_{p=i}^{k-1} \lambda_p \right) x_i \right].$$

Therefore, for random variables $X_1, X_2, \ldots, X_N$, we can define:

$$m_{N,\overrightarrow{\lambda}} = \sum_{i=1}^{N} \left[\left(\prod_{p=i}^{N-1} \lambda_p \right) X_i \right].$$

Similarly, it can be shown that $w_{N,\overrightarrow{\lambda}}$ can be expressed non-sequentially as:

$$w_{N,\overrightarrow{\lambda}} = \sum_{i=1}^{N} \left(\prod_{p=i}^{N-1} \lambda_p \right).$$

Now, the expectation of $\bar{X}_{N,\vec{\lambda}}$ can be computed by:

$$\begin{aligned}
\mathrm{E}\left[\bar{X}_{N,\vec{\lambda}}\right] &= \mathrm{E}\left[\frac{m_{N,\vec{\lambda}}}{w_{N,\vec{\lambda}}}\right] \\
&= \mathrm{E}\left[\frac{1}{w_{N,\vec{\lambda}}}\sum_{i=1}^{N}\left[\left(\prod_{p=i}^{N-1}\lambda_p\right)X_i\right]\right] \\
&= \frac{1}{w_{N,\vec{\lambda}}}\mathrm{E}\left[\sum_{i=1}^{N}\left[\left(\prod_{p=i}^{N-1}\lambda_p\right)X_i\right]\right] \\
&= \frac{1}{w_{N,\vec{\lambda}}}\sum_{i=1}^{N}\mathrm{E}\left[\left(\prod_{p=i}^{N-1}\lambda_p\right)X_i\right] \\
&= \frac{1}{w_{N,\vec{\lambda}}}\sum_{i=1}^{N}\left(\prod_{p=i}^{N-1}\lambda_p\right)\mathrm{E}\left[X_i\right] \\
&= \frac{1}{w_{N,\vec{\lambda}}}\sum_{i=1}^{N}\left(\prod_{p=i}^{N-1}\lambda_p\right)\mu \\
&= \left[\frac{1}{w_{N,\vec{\lambda}}}\sum_{i=1}^{N}\left(\prod_{p=i}^{N-1}\lambda_p\right)\right]\mu \\
\Rightarrow \mathrm{E}\left[\bar{X}_{N,\vec{\lambda}}\right] &= \mu.
\end{aligned}$$

The variance of $\bar{X}_{N,\vec{\lambda}}$ is computed by:

$$\begin{aligned}
\mathrm{Var}\left[\bar{X}_{N,\vec{\lambda}}\right] &= \mathrm{Var}\left[\frac{1}{w_{N,\vec{\lambda}}}\sum_{i=1}^{N}\left[\left(\prod_{p=i}^{N-1}\lambda_p\right)X_i\right]\right] \\
&= \frac{1}{(w_{N,\vec{\lambda}})^2}\mathrm{Var}\left[\sum_{i=1}^{N}\left[\left(\prod_{p=i}^{N-1}\lambda_p\right)X_i\right]\right]
\end{aligned}$$

and using the independence assumption:

$$\begin{aligned}
\mathrm{Var}\left[\bar{X}_{N,\vec{\lambda}}\right] &= \frac{1}{(w_{N,\vec{\lambda}})^2}\sum_{i=1}^{N}\mathrm{Var}\left[\left(\prod_{p=i}^{N-1}\lambda_p\right)X_i\right] \\
&= \frac{1}{(w_{N,\vec{\lambda}})^2}\sum_{i=1}^{N}\left(\prod_{p=i}^{N-1}\lambda_p\right)^2\mathrm{Var}\left[X_i\right] \\
&= \frac{1}{(w_{N,\vec{\lambda}})^2}\sum_{i=1}^{N}\left(\prod_{p=i}^{N-1}\lambda_p^2\right)\sigma^2
\end{aligned} \tag{17}$$

and by defining

$$u_{N,\vec{\lambda}} = \frac{1}{(w_{N,\vec{\lambda}})^2}\sum_{i=1}^{N}\left(\prod_{p=i}^{N-1}\lambda_p^2\right),$$

we finally have

$$\mathrm{Var}\left[\bar{X}_{N,\vec{\lambda}}\right] = \left(u_{N,\vec{\lambda}}\right)\sigma^2.$$

Part (b): All that remains is to derive the sequential update equation for $u_{N,\vec{\lambda}}$. We start with the case of $N = 1$. Recalling that $w_{0,\vec{\lambda}} = 0$, we first derive:

$$w_{1,\vec{\lambda}} = \lambda_1 w_{0,\vec{\lambda}} + 1 = \lambda_1 \cdot 0 + 1 = 1,$$

which makes sense, since for one observation we should have $w_{1,\vec{\lambda}} = 1$ for any value of $\vec{\lambda}$. Now, recalling that the empty product is defined to be 1,

$$\begin{aligned}
u_{1,\vec{\lambda}} &= \frac{1}{(w_{1,\vec{\lambda}})^2} \sum_{i=1}^{1} \left(\prod_{p=i}^{1-1} \lambda_p^2 \right) \\
&= \frac{1}{(1)^2} \left(\prod_{p=1}^{0} \lambda_p^2 \right) \\
\Rightarrow u_{1,\vec{\lambda}} &= 1.
\end{aligned}$$

Since,

$$u_{N,\vec{\lambda}} = \left(\frac{1}{w_{N,\vec{\lambda}}} \right)^2 \sum_{i=1}^{N} \Big(\prod_{p=i}^{N-1} (\lambda_p)^2 \Big), \tag{18}$$

we define the quantity $\tilde{u}_{N,\vec{\lambda}}$,

$$\begin{aligned}
\tilde{u}_{N,\vec{\lambda}} &= \sum_{i=1}^{N} \Big(\prod_{p=i}^{N-1} (\lambda_p)^2 \Big), \\
\Rightarrow \tilde{u}_{N,\vec{\lambda}} &= u_{N,\vec{\lambda}} (w_{N,\vec{\lambda}})^2.
\end{aligned}$$

Now consider $\tilde{u}_{N+1,\vec{\lambda}}$:

$$\begin{aligned}
\tilde{u}_{N+1,\vec{\lambda}} &= \sum_{i=1}^{N+1} \Big(\prod_{p=i}^{N} (\lambda_p)^2 \Big) \\
&= \sum_{i=1}^{N} \Big(\prod_{p=i}^{N} (\lambda_p)^2 \Big) + \Big(\prod_{p=N+1}^{N} (\lambda_p)^2 \Big) \\
&= \sum_{i=1}^{N} \Big(\prod_{p=i}^{N} (\lambda_p)^2 \Big) + 1 \\
&= (\lambda_N)^2 \sum_{i=1}^{N} \Big(\prod_{p=i}^{N-1} (\lambda_p)^2 \Big) + 1 \\
&= (\lambda_N)^2 \tilde{u}_{N,\vec{\lambda}} + 1.
\end{aligned}$$

This helps us with the sequential update equation for $\tilde{u}_{N+1,\overrightarrow{\lambda}}$:

$$\begin{aligned}
u_{N+1,\overrightarrow{\lambda}} &= \left(\frac{1}{w_{N+1,\overrightarrow{\lambda}}}\right)^2 \tilde{u}_{N+1,\overrightarrow{\lambda}} \\
&= \left(\frac{1}{w_{N+1,\overrightarrow{\lambda}}}\right)^2 \left[(\lambda_N)^2 \tilde{u}_{N,\overrightarrow{\lambda}} + 1\right] \\
&= \left(\frac{1}{w_{N+1,\overrightarrow{\lambda}}}\right)^2 \left[(\lambda_N)^2 u_{N,\overrightarrow{\lambda}} (w_{N,\overrightarrow{\lambda}})^2 + 1\right] \\
&= \left(\frac{\lambda_N w_{N,\overrightarrow{\lambda}}}{w_{N+1,\overrightarrow{\lambda}}}\right)^2 u_{N,\overrightarrow{\lambda}} + \left(\frac{1}{w_{N+1,\overrightarrow{\lambda}}}\right)^2 \\
&= \left(\frac{w_{N+1,\overrightarrow{\lambda}} - 1}{w_{N+1,\overrightarrow{\lambda}}}\right)^2 u_{N,\overrightarrow{\lambda}} + \left(\frac{1}{w_{N+1,\overrightarrow{\lambda}}}\right)^2 \\
\Rightarrow u_{N+1,\overrightarrow{\lambda}} &= \left(1 - \frac{1}{w_{N+1,\overrightarrow{\lambda}}}\right)^2 u_{N,\overrightarrow{\lambda}} + \left(\frac{1}{w_{N+1,\overrightarrow{\lambda}}}\right)^2
\end{aligned}$$

This completes the proof. $\square$

2.2 Expectation of cost function $L_{N,\overrightarrow{\lambda}}$

In this section we prove the proposition:

Proposition 7. *If our data stream is sampled from the i.i.d. random variables $X_1, X_2, \ldots, X_N$, with $E[X_i] = \mu$ and $Var[X_i] = \sigma^2$ for all $i \geq 1$, then*

$$E\left[\frac{\partial}{\partial \overrightarrow{\lambda}} L_{N,\overrightarrow{\lambda}}\right] \sim O(\sigma^2),$$

where $L_{N,\overrightarrow{\lambda}}$ is defined in Equation (18) *in Section* 4.2.2.

Proof:

Recall the definition of the cost function $L_{N,\overrightarrow{\lambda}}$ from Equation (18) in Section 4.2.2:

$$L_{N,\overrightarrow{\lambda}} = \left[\bar{x}_{N-1,\overrightarrow{\lambda}} - x_N\right]^2.$$

Although it was originally defined in terms of observations $x_1, x_2, \ldots, x_N$, we can similarly define $L_{N,\overrightarrow{\lambda}}$ in terms of the the random variables $X_1, X_2, \ldots X_N$:

$$L_{N,\overrightarrow{\lambda}} = \left[X_N - \bar{X}_{N-1,\overrightarrow{\lambda}}\right]^2.$$

Note we have switched the order of the terms inside the brackets, but this makes no difference to the value of the quantity. If we again assume that $\mathrm{E}[X_i] = \mu$ and $\mathrm{Var}[X_i] = \sigma^2$ for i.i.d. $X_i, X_2, \ldots, X_N$, then since X_N and $\bar{X}_{N-1,\overrightarrow{\lambda}}$ are independent, the expectation of the cost function $L_{N,\overrightarrow{\lambda}}$ is:

$$\begin{aligned}
\mathrm{E}\left[L_{N,\overrightarrow{\lambda}}\right] &= \mathrm{E}\left[\left[X_N - \bar{X}_{N-1,\overrightarrow{\lambda}}\right]^2\right] \\
&= \mathrm{E}\left[X_N^2 - 2X_N\bar{X}_{N-1,\overrightarrow{\lambda}} + \left(\bar{X}_{N-1,\overrightarrow{\lambda}}\right)^2\right] \\
&= \mathrm{E}\left[X_N^2\right] - 2\mathrm{E}\left[X_N\bar{X}_{N-1,\overrightarrow{\lambda}}\right] + \mathrm{E}\left[\left(\bar{X}_{N-1,\overrightarrow{\lambda}}\right)^2\right] \\
&= \left(\mathrm{Var}[X_N] + (\mathrm{E}[X_N])^2\right) - 2\mathrm{E}\left[X_N\right]\mathrm{E}\left[\bar{X}_{N-1,\overrightarrow{\lambda}}\right] + \\
&\qquad \left(\mathrm{Var}[\bar{X}_{N-1,\overrightarrow{\lambda}}] + \left(\mathrm{E}[\bar{X}_{N-1,\overrightarrow{\lambda}}]\right)^2\right) \\
&= (\sigma^2 + \mu^2) - 2\mu\cdot\mu + \left(u_{N-1,\overrightarrow{\lambda}}\sigma^2 + \mu^2\right) \\
\Rightarrow \mathrm{E}\left[L_{N,\overrightarrow{\lambda}}\right] &= \sigma^2\left(u_{N-1,\overrightarrow{\lambda}} + 1\right) \qquad (19)
\end{aligned}$$

which shows that $\mathrm{E}\left[L_{N,\overrightarrow{\lambda}}\right] \sim O(\sigma^2)$.

We now compute the expectation of $\mathrm{E}\left[\frac{\partial}{\partial \overrightarrow{\lambda}} L_{N+1,\overrightarrow{\lambda}}\right]$:

$$\begin{aligned}
\mathrm{E}\left[\frac{\partial}{\partial \overrightarrow{\lambda}} L_{N+1,\overrightarrow{\lambda}}\right] &= \mathrm{E}\left[\lim_{\epsilon\to 0}\frac{1}{\epsilon}\left(L_{N+1,\overrightarrow{\lambda}+\epsilon} - L_{N+1,\overrightarrow{\lambda}}\right)\right] \\
&= \lim_{\epsilon\to 0}\frac{1}{\epsilon}\left(\mathrm{E}\left[\ L_{N+1,\overrightarrow{\lambda}+\epsilon}\right] - \mathrm{E}\left[L_{N+1,\overrightarrow{\lambda}}\right]\right) \\
&= \lim_{\epsilon\to 0}\frac{1}{\epsilon}\left(\left[u_{N,\overrightarrow{\lambda}+\epsilon} + 1\right]\sigma^2 - \left[u_{N,\overrightarrow{\lambda}} + 1\right]\sigma^2\right) \\
&= \left[\lim_{\epsilon\to 0}\frac{1}{\epsilon}\left(u_{N,\overrightarrow{\lambda}+\epsilon} - u_{N,\overrightarrow{\lambda}}\right)\right]\sigma^2 \\
&= \left[\frac{\partial}{\partial \overrightarrow{\lambda}} u_{N,\overrightarrow{\lambda}}\right]\sigma^2
\end{aligned}$$

This is enough to show that $\mathrm{E}\left[\frac{\partial}{\partial \overrightarrow{\lambda}} L_{N+1,\overrightarrow{\lambda}}\right] \sim O(\sigma^2)$, which is sufficient for our purposes. In order to compute the value of this quantity exactly, in terms of $\overrightarrow{\lambda}$, we can proceed as follows:

$$
\begin{aligned}
\frac{\partial}{\partial \overrightarrow{\lambda}} u_{N,\overrightarrow{\lambda}} &= \frac{\partial}{\partial \overrightarrow{\lambda}} \left[\left(\frac{1}{w_{N,\overrightarrow{\lambda}}} \right)^2 \sum_{k=1}^{N} \left(\prod_{p=k}^{N-1} (\lambda_p)^2 \right) \right] \\
&= \frac{\partial}{\partial \overrightarrow{\lambda}} \left[\left(\frac{1}{w_{N,\overrightarrow{\lambda}}} \right)^2 \left(w_{N,\overrightarrow{\lambda}^2} \right) \right] \\
&= \left[\frac{\partial}{\partial \overrightarrow{\lambda}} \left(\frac{1}{w_{N,\overrightarrow{\lambda}}} \right)^2 \right] \left(w_{N,\overrightarrow{\lambda}^2} \right) + \left(\frac{1}{w_{N,\overrightarrow{\lambda}}} \right)^2 \left[\frac{\partial}{\partial \overrightarrow{\lambda}} \left(w_{N,\overrightarrow{\lambda}^2} \right) \right] \\
&= \left[\frac{-2}{(w_{N,\overrightarrow{\lambda}})^3} \Omega_{N,\overrightarrow{\lambda}} \right] \left(w_{N,\overrightarrow{\lambda}^2} \right) + \left(\frac{1}{w_{N,\overrightarrow{\lambda}}} \right)^2 \left[\frac{\partial}{\partial \overrightarrow{\lambda}} \left(w_{N,\overrightarrow{\lambda}^2} \right) \right] \\
&= \left[\frac{-2}{w_{N,\overrightarrow{\lambda}}} \Omega_{N,\overrightarrow{\lambda}} \frac{w_{N,\overrightarrow{\lambda}^2}}{(w_{N,\overrightarrow{\lambda}})^2} \right] + \left(\frac{1}{w_{N,\overrightarrow{\lambda}}} \right)^2 \left[\frac{\partial}{\partial \overrightarrow{\lambda}} \left(w_{N,\overrightarrow{\lambda}^2} \right) \right] \\
&= \frac{-2}{w_{N,\overrightarrow{\lambda}}} \Omega_{N,\overrightarrow{\lambda}} u_{N,\overrightarrow{\lambda}} + \left(\frac{1}{w_{N,\overrightarrow{\lambda}}} \right)^2 \left[\frac{\partial}{\partial \overrightarrow{\lambda}} \left(w_{N,\overrightarrow{\lambda}^2} \right) \right]
\end{aligned}
$$

Before computing the derivative of $w_{N,\overrightarrow{\lambda}^2}$, consider the following term:

$$
\left(\prod_{p=k}^{N-1} (\lambda_p + \epsilon)^2 \right) = \left(\prod_{p=k}^{N-1} \left(\lambda_p^2 + 2\lambda_p \epsilon + \epsilon^2 \right) \right) = \left(\prod_{t=k}^{N-1} \left(\lambda_t^2 + \epsilon(2\lambda_t + \epsilon) \right) \right)
$$

Using Lemma 4*, with $a_t = \lambda_t^2$ and $b_t = 2\lambda_t + \epsilon$, we have,

$$
\begin{aligned}
\left(\prod_{t=k}^{N-1} \left(\lambda_t^2 + \epsilon(2\lambda_t + \epsilon) \right) \right) &= \prod_{t=k}^{N-1} \lambda_t^2 + \epsilon \sum_{t=k}^{N-1} \left((2\lambda_t + \epsilon) \prod_{\substack{p=k \\ p \neq t}}^{N-1} \lambda_p^2 \right) + O(\epsilon^2) \\
&= \prod_{t=k}^{N-1} \lambda_t^2 + \epsilon \sum_{t=k}^{N-1} \left(2\lambda_t \prod_{\substack{p=k \\ p \neq t}}^{N-1} \lambda_p^2 \right) + O(\epsilon^2)
\end{aligned}
$$

We now easily compute the derivative of $w_{N,\overrightarrow{\lambda}^2}$:

$$
\begin{aligned}
\frac{\partial}{\partial \overrightarrow{\lambda}} \left(w_{N,\overrightarrow{\lambda}^2} \right) &= \frac{\partial}{\partial \overrightarrow{\lambda}} \left[\sum_{k=1}^{N} \left(\prod_{p=k}^{N-1} (\lambda_p)^2 \right) \right] \\
&= \lim_{\epsilon \to 0} \frac{1}{\epsilon} \left[\sum_{k=1}^{N} \left(\prod_{p=k}^{N-1} (\lambda_p + \epsilon)^2 \right) - \sum_{k=1}^{N} \left(\prod_{p=k}^{N-1} (\lambda_p)^2 \right) \right] \\
&= \lim_{\epsilon \to 0} \frac{1}{\epsilon} \left[\sum_{k=1}^{N} \left(\prod_{t=k}^{N-1} \lambda_t^2 + \epsilon \sum_{t=k}^{N-1} \left(2\lambda_t \prod_{\substack{p=k \\ p \neq t}}^{N-1} \lambda_p^2 \right) + O(\epsilon^2) \right) - \sum_{k=1}^{N} \left(\prod_{p=k}^{N-1} (\lambda_p)^2 \right) \right] \\
&= \lim_{\epsilon \to 0} \frac{1}{\epsilon} \left[\sum_{k=1}^{N} \left(\prod_{t=k}^{N-1} \lambda_t^2 \right) + \epsilon \sum_{k=1}^{N} \left(\sum_{t=k}^{N-1} \left(2\lambda_t \prod_{\substack{p=k \\ p \neq t}}^{N-1} \lambda_p^2 \right) \right) \right. \\
&\qquad \left. - \sum_{k=1}^{N} \left(\prod_{p=k}^{N-1} (\lambda_p)^2 \right) + O(\epsilon^2) \right]
\end{aligned}
$$

And cancelling terms, this leaves us with:

$$
\begin{aligned}
&= \lim_{\epsilon \to 0} \frac{1}{\epsilon} \left[\epsilon \sum_{k=1}^{N} \sum_{t=k}^{N-1} \left(2\lambda_t \prod_{\substack{p=k \\ p \neq t}}^{N-1} \lambda_p^2 \right) + O(\epsilon^2) \right] \\
&= 2 \lim_{\epsilon \to 0} \left[\sum_{k=1}^{N} \sum_{t=k}^{N-1} \left(\lambda_t \prod_{\substack{p=k \\ p \neq t}}^{N-1} \lambda_p^2 \right) + O(\epsilon) \right] \\
&= 2 \sum_{k=1}^{N} \sum_{t=k}^{N-1} \left(\lambda_t \prod_{\substack{p=k \\ p \neq t}}^{N-1} \lambda_p^2 \right) \\
&= 2 \sum_{k=1}^{N} \sum_{t=k}^{N-1} \lambda_t \left(\prod_{\substack{p=k \\ p \neq t}}^{N-1} \lambda_p \right)^2 \\
&= 2 \sum_{k=1}^{N} \sum_{t=k}^{N-1} \lambda_t \left(\prod_{\substack{q=k \\ q \neq t}}^{N-1} \lambda_q \right) \left(\prod_{\substack{p=k \\ p \neq t}}^{N-1} \lambda_p \right) \\
&= 2 \sum_{k=1}^{N} \sum_{t=k}^{N-1} \left(\prod_{q=k}^{N-1} \lambda_q \right) \left(\prod_{\substack{p=k \\ p \neq t}}^{N-1} \lambda_p \right) \\
&= 2 \sum_{k=1}^{N-1} \sum_{t=k}^{N-1} \left(\prod_{q=k}^{N-1} \lambda_q \right) \left(\prod_{\substack{p=k \\ p \neq t}}^{N-1} \lambda_p \right)
\end{aligned}
$$

where the last equality follows from the fact that when $k = N$, the second summation is zero. If we define the quantity,

$$
\psi_{N,\overrightarrow{\lambda}} = \sum_{k=1}^{N-1} \sum_{t=k}^{N-1} \left(\prod_{q=k}^{N-1} \lambda_q \right) \left(\prod_{\substack{p=k \\ p \neq t}}^{N-1} \lambda_p \right),
$$

we can write

$$
\frac{\partial}{\partial \overrightarrow{\lambda}} \left(w_{N,\overrightarrow{\lambda^2}} \right) = 2\psi_{N,\overrightarrow{\lambda}}.
$$

And so finally, we have the exact expression

$$
\begin{aligned}
\mathrm{E} \left[\frac{\partial}{\partial \overrightarrow{\lambda}} L_{N+1,\overrightarrow{\lambda}} \right] &= \left[\frac{\partial}{\partial \overrightarrow{\lambda}} u_{N,\overrightarrow{\lambda}} \right] \sigma^2 \\
&= \left[\frac{-2}{w_{N,\overrightarrow{\lambda}}} \Omega_{N,\overrightarrow{\lambda}} u_{N,\overrightarrow{\lambda}} + \left(\frac{1}{w_{N,\overrightarrow{\lambda}}} \right)^2 \left[\frac{\partial}{\partial \overrightarrow{\lambda}} \left(w_{N,\overrightarrow{\lambda^2}} \right) \right] \right] \sigma^2 \\
&= \left[\frac{-2}{w_{N,\overrightarrow{\lambda}}} \Omega_{N,\overrightarrow{\lambda}} u_{N,\overrightarrow{\lambda}} + \left(\frac{1}{w_{N,\overrightarrow{\lambda}}} \right)^2 2\psi_{N,\overrightarrow{\lambda}} \right] \sigma^2 \\
&= \frac{2}{w_{N,\overrightarrow{\lambda}}} \left(\frac{1}{w_{N,\overrightarrow{\lambda}}} \psi_{N,\overrightarrow{\lambda}} - \Omega_{N,\overrightarrow{\lambda}} u_{N,\lambda} \right) \sigma^2 \qquad (20)
\end{aligned}
$$

This completes the proof. This also agrees with an alternative proof in Bodenham (2014, Sec. A.3.9). □

3 Additional Tables

This section provides additional tables, similar to Table 2 in the main paper. In these tables, the step-size η is specified for the AFF algorithm in order to further demonstrate that the performance is similar for different η (but fixed α). This is specifically demonstrated in Section 3.3. Note that, as in the simulation study in the main text, each stream is generated as $\mathrm{N}(\mu, \sigma^2)$, where μ varies but $\sigma = 1$.

3.1 Tables of each algorithm as δ varies

In this section the tables are arranged by (a) method and (b) parameter pair, where δ is varied in each table. To be clear, each row of a table represents the performance over a stream with change sizes of magnitude δ. For example, the first row of Table 1 shows the performance of CUSUM with $(k,\ h)$ = (0.25, 8.01) for a stream with changes only of size $\delta = 0.25$, while the second row shows the performance on a stream with changes only of size $\delta = 0.5$.

Algo	Params	Values	δ	CCD	DNF	ARL1	(ARL1)	ARL0	(ARL0)
CUSUM	$(k,\ h)$	(0.25, 8.01)	(0.25)	0.72	0.84	50.53	(36.95)	281.57	(431.99)
CUSUM	$(k,\ h)$	(0.25, 8.01)	(0.50)	0.91	0.79	32.11	(26.32)	281.57	(431.99)
CUSUM	$(k,\ h)$	(0.25, 8.01)	(0.75)	0.95	0.75	20.06	(16.60)	281.57	(431.99)
CUSUM	$(k,\ h)$	(0.25, 8.01)	(1.00)	0.96	0.74	14.62	(13.51)	281.57	(431.99)
CUSUM	$(k,\ h)$	(0.25, 8.01)	(1.50)	0.97	0.74	10.33	(11.20)	281.57	(431.99)
CUSUM	$(k,\ h)$	(0.25, 8.01)	(2.00)	0.98	0.73	8.24	(9.99)	281.57	(431.99)
CUSUM	$(k,\ h)$	(0.25, 8.01)	(3.00)	0.98	0.73	6.97	(10.59)	281.57	(431.99)
CUSUM	$(k,\ h)$	(0.25, 8.01)	(4.00)	0.98	0.73	6.22	(10.06)	281.57	(431.99)

Table 1: This table shows the results for CUSUM with parameters $(k,\ h)$ = (0.25, 8.01) for streams with approximately 5000 changepoints and different δ values, using burn-in length B=50.

Algo	Params	Values	δ	CCD	DNF	ARL1	(ARL1)	ARL0	(ARL0)
CUSUM	$(k,\ h)$	(0.50, 4.77)	(0.25)	0.60	0.84	52.96	(39.19)	357.41	(582.96)
CUSUM	$(k,\ h)$	(0.50, 4.77)	(0.50)	0.84	0.79	36.32	(31.83)	357.41	(582.96)
CUSUM	$(k,\ h)$	(0.50, 4.77)	(0.75)	0.94	0.77	21.83	(21.32)	357.41	(582.96)
CUSUM	$(k,\ h)$	(0.50, 4.77)	(1.00)	0.95	0.75	14.38	(15.07)	357.41	(582.96)
CUSUM	$(k,\ h)$	(0.50, 4.77)	(1.50)	0.96	0.74	8.97	(11.57)	357.41	(582.96)
CUSUM	$(k,\ h)$	(0.50, 4.77)	(2.00)	0.97	0.74	7.05	(10.74)	357.41	(582.96)
CUSUM	$(k,\ h)$	(0.50, 4.77)	(3.00)	0.97	0.74	5.00	(7.77)	357.41	(582.96)
CUSUM	$(k,\ h)$	(0.50, 4.77)	(4.00)	0.96	0.74	4.41	(7.54)	357.41	(582.96)

Table 2: This table shows the results for CUSUM with parameters $(k,\ h)$ = (0.50, 4.77) for streams with approximately 5000 changepoints and different δ values, using burn-in length B=50.

Algo	Params	Values	δ	CCD	DNF	ARL1	(ARL1)	ARL0	(ARL0)
CUSUM	$(k,\ h)$	(0.75, 3.34)	(0.25)	0.51	0.85	54.60	(40.52)	440.62	(757.59)
CUSUM	$(k,\ h)$	(0.75, 3.34)	(0.50)	0.75	0.82	40.73	(35.43)	440.62	(757.59)
CUSUM	$(k,\ h)$	(0.75, 3.34)	(0.75)	0.91	0.79	26.81	(27.03)	440.62	(757.59)
CUSUM	$(k,\ h)$	(0.75, 3.34)	(1.00)	0.94	0.76	16.44	(18.06)	440.62	(757.59)
CUSUM	$(k,\ h)$	(0.75, 3.34)	(1.50)	0.95	0.75	8.92	(12.33)	440.62	(757.59)
CUSUM	$(k,\ h)$	(0.75, 3.34)	(2.00)	0.95	0.75	6.61	(10.80)	440.62	(757.59)
CUSUM	$(k,\ h)$	(0.75, 3.34)	(3.00)	0.95	0.75	4.82	(9.36)	440.62	(757.59)
CUSUM	$(k,\ h)$	(0.75, 3.34)	(4.00)	0.95	0.74	3.95	(7.84)	440.62	(757.59)

Table 3: This table shows the results for CUSUM with parameters $(k,\ h)$ = (0.75, 3.34) for streams with approximately 5000 changepoints and different δ values, using burn-in length B=50.

Algo	Params	Values	δ	CCD	DNF	ARL1	(ARL1)	ARL0	(ARL0)
CUSUM	(k, h)	(1.00, 2.52)	(0.25)	0.44	0.86	57.40	(41.14)	479.24	(848.11)
CUSUM	(k, h)	(1.00, 2.52)	(0.50)	0.67	0.83	45.18	(37.59)	479.24	(848.11)
CUSUM	(k, h)	(1.00, 2.52)	(0.75)	0.84	0.80	31.72	(30.80)	479.24	(848.11)
CUSUM	(k, h)	(1.00, 2.52)	(1.00)	0.92	0.78	20.20	(22.32)	479.24	(848.11)
CUSUM	(k, h)	(1.00, 2.52)	(1.50)	0.93	0.76	10.04	(14.48)	479.24	(848.11)
CUSUM	(k, h)	(1.00, 2.52)	(2.00)	0.93	0.75	6.74	(11.86)	479.24	(848.11)
CUSUM	(k, h)	(1.00, 2.52)	(3.00)	0.93	0.75	4.59	(9.48)	479.24	(848.11)
CUSUM	(k, h)	(1.00, 2.52)	(4.00)	0.93	0.75	3.57	(7.70)	479.24	(848.11)

Table 4: This table shows the results for CUSUM with parameters $(k, h) = (1.00, 2.52)$ for streams with approximately 5000 changepoints and different δ values, using burn-in length B=50.

Algo	Params	Values	δ	CCD	DNF	ARL1	(ARL1)	ARL0	(ARL0)
CUSUM	(k, h)	(1.25, 1.99)	(0.25)	0.39	0.86	58.89	(41.88)	526.88	(906.81)
CUSUM	(k, h)	(1.25, 1.99)	(0.50)	0.61	0.84	48.16	(39.13)	526.88	(906.81)
CUSUM	(k, h)	(1.25, 1.99)	(0.75)	0.78	0.81	36.58	(34.19)	526.88	(906.81)
CUSUM	(k, h)	(1.25, 1.99)	(1.00)	0.88	0.79	25.58	(27.33)	526.88	(906.81)
CUSUM	(k, h)	(1.25, 1.99)	(1.50)	0.91	0.76	11.50	(15.92)	526.88	(906.81)
CUSUM	(k, h)	(1.25, 1.99)	(2.00)	0.91	0.75	7.19	(12.90)	526.88	(906.81)
CUSUM	(k, h)	(1.25, 1.99)	(3.00)	0.91	0.75	4.60	(10.53)	526.88	(906.81)
CUSUM	(k, h)	(1.25, 1.99)	(4.00)	0.90	0.75	3.58	(8.71)	526.88	(906.81)

Table 5: This table shows the results for CUSUM with parameters $(k, h) = (1.25, 1.99)$ for streams with approximately 5000 changepoints and different δ values, using burn-in length B=50.

Algo	Params	Values	δ	CCD	DNF	ARL1	(ARL1)	ARL0	(ARL0)
CUSUM	(k, h)	(1.50, 1.61)	(0.25)	0.36	0.86	60.24	(41.82)	493.38	(855.65)
CUSUM	(k, h)	(1.50, 1.61)	(0.50)	0.56	0.85	51.10	(39.99)	493.38	(855.65)
CUSUM	(k, h)	(1.50, 1.61)	(0.75)	0.71	0.82	40.66	(36.66)	493.38	(855.65)
CUSUM	(k, h)	(1.50, 1.61)	(1.00)	0.83	0.80	30.37	(31.35)	493.38	(855.65)
CUSUM	(k, h)	(1.50, 1.61)	(1.50)	0.90	0.77	13.97	(18.66)	493.38	(855.65)
CUSUM	(k, h)	(1.50, 1.61)	(2.00)	0.89	0.76	8.03	(14.35)	493.38	(855.65)
CUSUM	(k, h)	(1.50, 1.61)	(3.00)	0.88	0.75	4.39	(10.19)	493.38	(855.65)
CUSUM	(k, h)	(1.50, 1.61)	(4.00)	0.87	0.75	3.12	(6.88)	493.38	(855.65)

Table 6: This table shows the results for CUSUM with parameters $(k, h) = (1.50, 1.61)$ for streams with approximately 5000 changepoints and different δ values, using burn-in length B=50.

Algo	Params	Values	δ	CCD	DNF	ARL1	(ARL1)	ARL0	(ARL0)
EWMA	(r, L)	(1.00, 3.090)	(0.25)	0.29	0.90	64.82	(43.04)	661.37	(987.72)
EWMA	(r, L)	(1.00, 3.090)	(0.50)	0.44	0.88	56.59	(41.03)	661.37	(987.72)
EWMA	(r, L)	(1.00, 3.090)	(0.75)	0.56	0.87	48.72	(39.34)	661.37	(987.72)
EWMA	(r, L)	(1.00, 3.090)	(1.00)	0.68	0.86	42.14	(37.63)	661.37	(987.72)
EWMA	(r, L)	(1.00, 3.090)	(1.50)	0.85	0.83	25.31	(28.20)	661.37	(987.72)
EWMA	(r, L)	(1.00, 3.090)	(2.00)	0.87	0.80	12.79	(18.06)	661.37	(987.72)
EWMA	(r, L)	(1.00, 3.090)	(3.00)	0.87	0.79	5.23	(10.83)	661.37	(987.72)
EWMA	(r, L)	(1.00, 3.090)	(4.00)	0.86	0.79	3.78	(9.78)	661.37	(987.72)

Table 7: This table shows the results for EWMA with parameters $(r, L) = (1.00, 3.090)$ for streams with approximately 5000 changepoints and different δ values, using burn-in length B=50.

Algo	Params	Values	δ	CCD	DNF	ARL1	(ARL1)	ARL0	(ARL0)
EWMA	(r, L)	(0.75, 3.087)	(0.25)	0.33	0.88	62.23	(41.73)	676.26	(1072.18)
EWMA	(r, L)	(0.75, 3.087)	(0.50)	0.51	0.87	52.72	(40.34)	676.26	(1072.18)
EWMA	(r, L)	(0.75, 3.087)	(0.75)	0.67	0.85	42.66	(37.53)	676.26	(1072.18)
EWMA	(r, L)	(0.75, 3.087)	(1.00)	0.80	0.84	33.73	(33.26)	676.26	(1072.18)
EWMA	(r, L)	(0.75, 3.087)	(1.50)	0.91	0.80	16.12	(19.92)	676.26	(1072.18)
EWMA	(r, L)	(0.75, 3.087)	(2.00)	0.90	0.79	8.59	(13.56)	676.26	(1072.18)
EWMA	(r, L)	(0.75, 3.087)	(3.00)	0.88	0.78	4.56	(10.64)	676.26	(1072.18)
EWMA	(r, L)	(0.75, 3.087)	(4.00)	0.87	0.79	3.43	(9.17)	676.26	(1072.18)

Table 8: This table shows the results for EWMA with parameters $(r, L) = (0.75, 3.087)$ for streams with approximately 5000 changepoints and different δ values, using burn-in length B=50.

Algo	Params	Values	δ	CCD	DNF	ARL1	(ARL1)	ARL0	(ARL0)
EWMA	(r, L)	(0.50, 3.071)	(0.25)	0.39	0.88	60.23	(41.54)	638.20	(983.45)
EWMA	(r, L)	(0.50, 3.071)	(0.50)	0.62	0.85	47.67	(38.90)	638.20	(983.45)
EWMA	(r, L)	(0.50, 3.071)	(0.75)	0.80	0.84	36.97	(34.81)	638.20	(983.45)
EWMA	(r, L)	(0.50, 3.071)	(1.00)	0.90	0.81	24.34	(26.18)	638.20	(983.45)
EWMA	(r, L)	(0.50, 3.071)	(1.50)	0.93	0.79	10.89	(14.71)	638.20	(983.45)
EWMA	(r, L)	(0.50, 3.071)	(2.00)	0.92	0.78	7.08	(12.15)	638.20	(983.45)
EWMA	(r, L)	(0.50, 3.071)	(3.00)	0.91	0.78	4.52	(9.68)	638.20	(983.45)
EWMA	(r, L)	(0.50, 3.071)	(4.00)	0.90	0.78	3.61	(8.95)	638.20	(983.45)

Table 9: This table shows the results for EWMA with parameters $(r, L) = (0.50, 3.071)$ for streams with approximately 5000 changepoints and different δ values, using burn-in length B=50.

Algo	Params	Values	δ	CCD	DNF	ARL1	(ARL1)	ARL0	(ARL0)
EWMA	(r, L)	(0.40, 3.054)	(0.25)	0.44	0.87	57.75	(41.60)	564.67	(886.39)
EWMA	(r, L)	(0.40, 3.054)	(0.50)	0.67	0.85	45.55	(38.32)	564.67	(886.39)
EWMA	(r, L)	(0.40, 3.054)	(0.75)	0.85	0.82	32.87	(31.69)	564.67	(886.39)
EWMA	(r, L)	(0.40, 3.054)	(1.00)	0.92	0.80	20.13	(21.81)	564.67	(886.39)
EWMA	(r, L)	(0.40, 3.054)	(1.50)	0.94	0.79	9.82	(13.44)	564.67	(886.39)
EWMA	(r, L)	(0.40, 3.054)	(2.00)	0.94	0.78	6.82	(11.62)	564.67	(886.39)
EWMA	(r, L)	(0.40, 3.054)	(3.00)	0.93	0.78	4.90	(10.66)	564.67	(886.39)
EWMA	(r, L)	(0.40, 3.054)	(4.00)	0.92	0.78	3.73	(8.70)	564.67	(886.39)

Table 10: This table shows the results for EWMA with parameters $(r, L) = (0.40, 3.054)$ for streams with approximately 5000 changepoints and different δ values, using burn-in length B=50.

Algo	Params	Values	δ	CCD	DNF	ARL1	(ARL1)	ARL0	(ARL0)
EWMA	(r, L)	(0.30, 3.023)	(0.25)	0.47	0.87	57.02	(41.50)	506.76	(819.15)
EWMA	(r, L)	(0.30, 3.023)	(0.50)	0.74	0.84	42.11	(36.19)	506.76	(819.15)
EWMA	(r, L)	(0.30, 3.023)	(0.75)	0.90	0.81	28.43	(28.06)	506.76	(819.15)
EWMA	(r, L)	(0.30, 3.023)	(1.00)	0.94	0.79	17.26	(18.71)	506.76	(819.15)
EWMA	(r, L)	(0.30, 3.023)	(1.50)	0.95	0.78	9.09	(12.69)	506.76	(819.15)
EWMA	(r, L)	(0.30, 3.023)	(2.00)	0.95	0.78	6.49	(10.87)	506.76	(819.15)
EWMA	(r, L)	(0.30, 3.023)	(3.00)	0.95	0.77	4.57	(9.15)	506.76	(819.15)
EWMA	(r, L)	(0.30, 3.023)	(4.00)	0.94	0.77	3.81	(8.15)	506.76	(819.15)

Table 11: This table shows the results for EWMA with parameters $(r, L) = (0.30, 3.023)$ for streams with approximately 5000 changepoints and different δ values, using burn-in length B=50.

Algo	Params	Values	δ	CCD	DNF	ARL1	(ARL1)	ARL0	(ARL0)
EWMA	(r, L)	(0.25, 2.998)	(0.25)	0.51	0.87	55.79	(41.10)	502.91	(821.13)
EWMA	(r, L)	(0.25, 2.998)	(0.50)	0.77	0.83	40.81	(35.31)	502.91	(821.13)
EWMA	(r, L)	(0.25, 2.998)	(0.75)	0.91	0.80	26.26	(26.29)	502.91	(821.13)
EWMA	(r, L)	(0.25, 2.998)	(1.00)	0.94	0.79	16.14	(17.54)	502.91	(821.13)
EWMA	(r, L)	(0.25, 2.998)	(1.50)	0.96	0.78	8.42	(11.00)	502.91	(821.13)
EWMA	(r, L)	(0.25, 2.998)	(2.00)	0.96	0.77	6.17	(9.58)	502.91	(821.13)
EWMA	(r, L)	(0.25, 2.998)	(3.00)	0.95	0.77	4.53	(8.70)	502.91	(821.13)
EWMA	(r, L)	(0.25, 2.998)	(4.00)	0.95	0.77	3.65	(7.24)	502.91	(821.13)

Table 12: This table shows the results for EWMA with parameters $(r, L) = (0.25, 2.998)$ for streams with approximately 5000 changepoints and different δ values, using burn-in length B=50.

Algo	Params	Values	δ	CCD	DNF	ARL1	(ARL1)	ARL0	(ARL0)
EWMA	(r, L)	(0.20, 2.962)	(0.25)	0.55	0.86	54.43	(40.17)	493.38	(799.77)
EWMA	(r, L)	(0.20, 2.962)	(0.50)	0.81	0.83	38.43	(33.22)	493.38	(799.77)
EWMA	(r, L)	(0.20, 2.962)	(0.75)	0.93	0.80	23.96	(23.71)	493.38	(799.77)
EWMA	(r, L)	(0.20, 2.962)	(1.00)	0.95	0.78	14.73	(15.44)	493.38	(799.77)
EWMA	(r, L)	(0.20, 2.962)	(1.50)	0.96	0.77	8.03	(9.72)	493.38	(799.77)
EWMA	(r, L)	(0.20, 2.962)	(2.00)	0.96	0.77	6.03	(8.82)	493.38	(799.77)
EWMA	(r, L)	(0.20, 2.962)	(3.00)	0.96	0.77	4.47	(8.33)	493.38	(799.77)
EWMA	(r, L)	(0.20, 2.962)	(4.00)	0.96	0.76	3.60	(6.82)	493.38	(799.77)

Table 13: This table shows the results for EWMA with parameters $(r, L) = (0.20, 2.962)$ for streams with approximately 5000 changepoints and different δ values, using burn-in length B=50.

Algo	Params	Values	δ	CCD	DNF	ARL1	(ARL1)	ARL0	(ARL0)
EWMA	(r, L)	(0.10, 2.814)	(0.25)	0.64	0.86	52.87	(38.97)	407.02	(690.85)
EWMA	(r, L)	(0.10, 2.814)	(0.50)	0.89	0.81	34.29	(29.47)	407.02	(690.85)
EWMA	(r, L)	(0.10, 2.814)	(0.75)	0.95	0.77	19.82	(17.75)	407.02	(690.85)
EWMA	(r, L)	(0.10, 2.814)	(1.00)	0.96	0.75	13.06	(11.75)	407.02	(690.85)
EWMA	(r, L)	(0.10, 2.814)	(1.50)	0.97	0.71	8.49	(10.54)	407.02	(690.85)
EWMA	(r, L)	(0.10, 2.814)	(2.00)	0.97	0.66	6.42	(9.25)	407.02	(690.85)
EWMA	(r, L)	(0.10, 2.814)	(3.00)	0.97	0.56	5.18	(9.79)	407.02	(690.85)
EWMA	(r, L)	(0.10, 2.814)	(4.00)	0.97	0.49	4.61	(9.96)	407.02	(690.85)

Table 14: This table shows the results for EWMA with parameters $(r, L) = (0.10, 2.814)$ for streams with approximately 5000 changepoints and different δ values, using burn-in length B=50.

Algo	Params	Values	δ	CCD	DNF	ARL1	(ARL1)	ARL0	(ARL0)
EWMA	(r, L)	(0.05, 2.615)	(0.25)	0.87	0.41	17.23	(32.28)	356.06	(568.99)
EWMA	(r, L)	(0.05, 2.615)	(0.50)	0.98	0.36	5.82	(16.41)	356.06	(568.99)
EWMA	(r, L)	(0.05, 2.615)	(0.75)	0.99	0.35	2.71	(7.97)	356.06	(568.99)
EWMA	(r, L)	(0.05, 2.615)	(1.00)	1.00	0.35	2.06	(5.89)	356.06	(568.99)
EWMA	(r, L)	(0.05, 2.615)	(1.50)	1.00	0.35	1.59	(4.10)	356.06	(568.99)
EWMA	(r, L)	(0.05, 2.615)	(2.00)	1.00	0.35	1.41	(3.70)	356.06	(568.99)
EWMA	(r, L)	(0.05, 2.615)	(3.00)	1.00	0.34	1.23	(2.35)	356.06	(568.99)
EWMA	(r, L)	(0.05, 2.615)	(4.00)	1.00	0.34	1.20	(2.07)	356.06	(568.99)

Table 15: This table shows the results for EWMA with parameters $(r, L) = (0.05, 2.615)$ for streams with approximately 5000 changepoints and different δ values, using burn-in length B=50.

Algo	Params	Values	δ	CCD	DNF	ARL1	(ARL1)	ARL0	(ARL0)
EWMA	(r, L)	(0.03, 2.437)	(0.25)	0.99	0.35	2.89	(12.01)	312.45	(456.83)
EWMA	(r, L)	(0.03, 2.437)	(0.50)	1.00	0.34	1.91	(7.29)	312.45	(456.83)
EWMA	(r, L)	(0.03, 2.437)	(0.75)	1.00	0.34	1.39	(3.39)	312.45	(456.83)
EWMA	(r, L)	(0.03, 2.437)	(1.00)	1.00	0.34	1.30	(3.20)	312.45	(456.83)
EWMA	(r, L)	(0.03, 2.437)	(1.50)	1.00	0.34	1.19	(3.02)	312.45	(456.83)
EWMA	(r, L)	(0.03, 2.437)	(2.00)	1.00	0.34	1.14	(2.22)	312.45	(456.83)
EWMA	(r, L)	(0.03, 2.437)	(3.00)	1.00	0.34	1.11	(1.90)	312.45	(456.83)
EWMA	(r, L)	(0.03, 2.437)	(4.00)	1.00	0.34	1.06	(1.31)	312.45	(456.83)

Table 16: This table shows the results for EWMA with parameters $(r, L) = (0.03, 2.437)$ for streams with approximately 5000 changepoints and different δ values, using burn-in length B=50.

Algo	Params	Values	δ	CCD	DNF	ARL1	(ARL1)	ARL0	(ARL0)
AFF	(η, α)	(0.100, 0.005)	(0.25)	0.55	0.88	56.67	(39.80)	670.03	(972.68)
AFF	(η, α)	(0.100, 0.005)	(0.50)	0.86	0.84	40.61	(32.27)	670.03	(972.68)
AFF	(η, α)	(0.100, 0.005)	(0.75)	0.96	0.81	24.13	(20.46)	670.03	(972.68)
AFF	(η, α)	(0.100, 0.005)	(1.00)	0.97	0.80	15.24	(13.01)	670.03	(972.68)
AFF	(η, α)	(0.100, 0.005)	(1.50)	0.97	0.78	9.07	(9.76)	670.03	(972.68)
AFF	(η, α)	(0.100, 0.005)	(2.00)	0.98	0.78	7.06	(9.48)	670.03	(972.68)
AFF	(η, α)	(0.100, 0.005)	(3.00)	0.98	0.78	5.48	(8.76)	670.03	(972.68)
AFF	(η, α)	(0.100, 0.005)	(4.00)	0.98	0.77	5.30	(10.28)	670.03	(972.68)

Table 17: This table shows the results for AFF with parameters $(\eta, \alpha) = (0.100, 0.005)$ for streams with approximately 5000 changepoints and different δ values, using burn-in length B=50.

Algo	Params	Values	δ	CCD	DNF	ARL1	(ARL1)	ARL0	(ARL0)
AFF	(η, α)	(0.100, 0.006)	(0.25)	0.58	0.86	55.24	(39.68)	576.34	(893.73)
AFF	(η, α)	(0.100, 0.006)	(0.50)	0.86	0.83	39.28	(31.83)	576.34	(893.73)
AFF	(η, α)	(0.100, 0.006)	(0.75)	0.96	0.79	23.12	(19.85)	576.34	(893.73)
AFF	(η, α)	(0.100, 0.006)	(1.00)	0.97	0.78	14.89	(13.02)	576.34	(893.73)
AFF	(η, α)	(0.100, 0.006)	(1.50)	0.97	0.76	9.00	(10.08)	576.34	(893.73)
AFF	(η, α)	(0.100, 0.006)	(2.00)	0.98	0.76	6.98	(9.33)	576.34	(893.73)
AFF	(η, α)	(0.100, 0.006)	(3.00)	0.98	0.76	5.53	(9.07)	576.34	(893.73)
AFF	(η, α)	(0.100, 0.006)	(4.00)	0.98	0.75	5.34	(10.42)	576.34	(893.73)

Table 18: This table shows the results for AFF with parameters $(\eta, \alpha) = (0.100, 0.006)$ for streams with approximately 5000 changepoints and different δ values, using burn-in length B=50.

Algo	Params	Values	δ	CCD	DNF	ARL1	(ARL1)	ARL0	(ARL0)
AFF	(η, α)	(0.100, 0.007)	(0.25)	0.60	0.85	54.23	(39.64)	494.74	(765.15)
AFF	(η, α)	(0.100, 0.007)	(0.50)	0.88	0.81	38.08	(31.13)	494.74	(765.15)
AFF	(η, α)	(0.100, 0.007)	(0.75)	0.96	0.78	22.42	(19.41)	494.74	(765.15)
AFF	(η, α)	(0.100, 0.007)	(1.00)	0.97	0.76	14.41	(13.04)	494.74	(765.15)
AFF	(η, α)	(0.100, 0.007)	(1.50)	0.97	0.75	8.84	(10.12)	494.74	(765.15)
AFF	(η, α)	(0.100, 0.007)	(2.00)	0.98	0.74	7.09	(9.85)	494.74	(765.15)
AFF	(η, α)	(0.100, 0.007)	(3.00)	0.98	0.74	5.55	(9.07)	494.74	(765.15)
AFF	(η, α)	(0.100, 0.007)	(4.00)	0.98	0.74	5.40	(10.43)	494.74	(765.15)

Table 19: This table shows the results for AFF with parameters $(\eta, \alpha) = (0.100, 0.007)$ for streams with approximately 5000 changepoints and different δ values, using burn-in length B=50.

Algo	Params	Values	δ	CCD	DNF	ARL1	(ARL1)	ARL0	(ARL0)
AFF	(η, α)	(0.100, 0.008)	(0.25)	0.62	0.83	53.49	(39.69)	396.80	(629.73)
AFF	(η, α)	(0.100, 0.008)	(0.50)	0.89	0.79	37.21	(30.90)	396.80	(629.73)
AFF	(η, α)	(0.100, 0.008)	(0.75)	0.96	0.76	21.86	(19.18)	396.80	(629.73)
AFF	(η, α)	(0.100, 0.008)	(1.00)	0.96	0.75	14.10	(12.86)	396.80	(629.73)
AFF	(η, α)	(0.100, 0.008)	(1.50)	0.97	0.73	8.77	(10.32)	396.80	(629.73)
AFF	(η, α)	(0.100, 0.008)	(2.00)	0.97	0.73	7.03	(10.02)	396.80	(629.73)
AFF	(η, α)	(0.100, 0.008)	(3.00)	0.98	0.73	5.55	(9.01)	396.80	(629.73)
AFF	(η, α)	(0.100, 0.008)	(4.00)	0.98	0.72	5.50	(10.84)	396.80	(629.73)

Table 20: This table shows the results for AFF with parameters (η, α) = (0.100, 0.008) for streams with approximately 5000 changepoints and different δ values, using burn-in length B=50.

Algo	Params	Values	δ	CCD	DNF	ARL1	(ARL1)	ARL0	(ARL0)
AFF	(η, α)	(0.100, 0.009)	(0.25)	0.64	0.82	52.88	(39.13)	388.93	(634.65)
AFF	(η, α)	(0.100, 0.009)	(0.50)	0.89	0.78	36.22	(30.23)	388.93	(634.65)
AFF	(η, α)	(0.100, 0.009)	(0.75)	0.96	0.75	21.28	(18.78)	388.93	(634.65)
AFF	(η, α)	(0.100, 0.009)	(1.00)	0.97	0.73	13.94	(13.08)	388.93	(634.65)
AFF	(η, α)	(0.100, 0.009)	(1.50)	0.97	0.72	8.74	(10.65)	388.93	(634.65)
AFF	(η, α)	(0.100, 0.009)	(2.00)	0.97	0.72	7.02	(10.11)	388.93	(634.65)
AFF	(η, α)	(0.100, 0.009)	(3.00)	0.98	0.72	5.68	(9.53)	388.93	(634.65)
AFF	(η, α)	(0.100, 0.009)	(4.00)	0.98	0.71	5.71	(11.31)	388.93	(634.65)

Table 21: This table shows the results for AFF with parameters (η, α) = (0.100, 0.009) for streams with approximately 5000 changepoints and different δ values, using burn-in length B=50.

Algo	Params	Values	δ	CCD	DNF	ARL1	(ARL1)	ARL0	(ARL0)
AFF	(η, α)	(0.100, 0.010)	(0.25)	0.65	0.81	51.83	(38.99)	352.77	(557.18)
AFF	(η, α)	(0.100, 0.010)	(0.50)	0.90	0.76	35.16	(29.78)	352.77	(557.18)
AFF	(η, α)	(0.100, 0.010)	(0.75)	0.96	0.74	20.92	(18.71)	352.77	(557.18)
AFF	(η, α)	(0.100, 0.010)	(1.00)	0.96	0.72	13.79	(13.35)	352.77	(557.18)
AFF	(η, α)	(0.100, 0.010)	(1.50)	0.97	0.71	8.78	(10.84)	352.77	(557.18)
AFF	(η, α)	(0.100, 0.010)	(2.00)	0.97	0.70	7.00	(9.97)	352.77	(557.18)
AFF	(η, α)	(0.100, 0.010)	(3.00)	0.98	0.70	5.78	(9.81)	352.77	(557.18)
AFF	(η, α)	(0.100, 0.010)	(4.00)	0.98	0.70	5.80	(11.53)	352.77	(557.18)

Table 22: This table shows the results for AFF with parameters (η, α) = (0.100, 0.010) for streams with approximately 5000 changepoints and different δ values, using burn-in length B=50.

Algo	Params	Values	δ	CCD	DNF	ARL1	(ARL1)	ARL0	(ARL0)
AFF	(η, α)	(0.100, 0.025)	(0.25)	0.79	0.70	44.19	(36.70)	149.56	(199.33)
AFF	(η, α)	(0.100, 0.025)	(0.50)	0.93	0.65	28.11	(26.19)	149.56	(199.33)
AFF	(η, α)	(0.100, 0.025)	(0.75)	0.96	0.62	17.04	(17.22)	149.56	(199.33)
AFF	(η, α)	(0.100, 0.025)	(1.00)	0.97	0.61	12.37	(14.35)	149.56	(199.33)
AFF	(η, α)	(0.100, 0.025)	(1.50)	0.98	0.60	8.47	(12.14)	149.56	(199.33)
AFF	(η, α)	(0.100, 0.025)	(2.00)	0.97	0.60	6.88	(10.88)	149.56	(199.33)
AFF	(η, α)	(0.100, 0.025)	(3.00)	0.98	0.59	6.17	(11.53)	149.56	(199.33)
AFF	(η, α)	(0.100, 0.025)	(4.00)	0.98	0.59	5.96	(11.72)	149.56	(199.33)

Table 23: This table shows the results for AFF with parameters (η, α) = (0.100, 0.025) for streams with approximately 5000 changepoints and different δ values, using burn-in length B=50.

Algo	Params	Values	δ	CCD	DNF	ARL1	(ARL1)	ARL0	(ARL0)
AFF	(η, α)	(0.100, 0.050)	(0.25)	0.90	0.60	33.91	(32.19)	87.67	(95.61)
AFF	(η, α)	(0.100, 0.050)	(0.50)	0.96	0.56	22.16	(23.01)	87.67	(95.61)
AFF	(η, α)	(0.100, 0.050)	(0.75)	0.97	0.54	14.96	(17.53)	87.67	(95.61)
AFF	(η, α)	(0.100, 0.050)	(1.00)	0.98	0.53	11.31	(15.09)	87.67	(95.61)
AFF	(η, α)	(0.100, 0.050)	(1.50)	0.98	0.52	8.19	(13.03)	87.67	(95.61)
AFF	(η, α)	(0.100, 0.050)	(2.00)	0.98	0.52	6.58	(11.34)	87.67	(95.61)
AFF	(η, α)	(0.100, 0.050)	(3.00)	0.99	0.52	5.77	(10.99)	87.67	(95.61)
AFF	(η, α)	(0.100, 0.050)	(4.00)	0.98	0.51	5.58	(11.20)	87.67	(95.61)

Table 24: This table shows the results for AFF with parameters (η, α) = (0.100, 0.050) for streams with approximately 5000 changepoints and different δ values, using burn-in length B=50.

Algo	Params	Values	δ	CCD	DNF	ARL1	(ARL1)	ARL0	(ARL0)
AFF	(η, α)	(0.010, 0.005)	(0.25)	0.58	0.86	56.65	(40.34)	766.12	(1051.05)
AFF	(η, α)	(0.010, 0.005)	(0.50)	0.88	0.83	42.02	(32.92)	766.12	(1051.05)
AFF	(η, α)	(0.010, 0.005)	(0.75)	0.96	0.79	25.03	(21.12)	766.12	(1051.05)
AFF	(η, α)	(0.010, 0.005)	(1.00)	0.97	0.77	16.47	(14.63)	766.12	(1051.05)
AFF	(η, α)	(0.010, 0.005)	(1.50)	0.97	0.74	10.00	(11.35)	766.12	(1051.05)
AFF	(η, α)	(0.010, 0.005)	(2.00)	0.97	0.74	7.58	(9.99)	766.12	(1051.05)
AFF	(η, α)	(0.010, 0.005)	(3.00)	0.97	0.73	6.00	(10.57)	766.12	(1051.05)
AFF	(η, α)	(0.010, 0.005)	(4.00)	0.97	0.73	5.09	(10.13)	766.12	(1051.05)

Table 25: This table shows the results for AFF with parameters (η, α) = (0.010, 0.005) for streams with approximately 5000 changepoints and different δ values, using burn-in length B=50.

Algo	Params	Values	δ	CCD	DNF	ARL1	(ARL1)	ARL0	(ARL0)
AFF	(η, α)	(0.010, 0.006)	(0.25)	0.60	0.85	56.04	(40.08)	697.17	(1011.78)
AFF	(η, α)	(0.010, 0.006)	(0.50)	0.89	0.82	40.08	(32.36)	697.17	(1011.78)
AFF	(η, α)	(0.010, 0.006)	(0.75)	0.95	0.76	23.68	(19.94)	697.17	(1011.78)
AFF	(η, α)	(0.010, 0.006)	(1.00)	0.96	0.74	15.95	(14.66)	697.17	(1011.78)
AFF	(η, α)	(0.010, 0.006)	(1.50)	0.97	0.72	9.73	(11.03)	697.17	(1011.78)
AFF	(η, α)	(0.010, 0.006)	(2.00)	0.97	0.72	7.85	(11.30)	697.17	(1011.78)
AFF	(η, α)	(0.010, 0.006)	(3.00)	0.97	0.72	6.10	(11.02)	697.17	(1011.78)
AFF	(η, α)	(0.010, 0.006)	(4.00)	0.97	0.71	5.24	(10.86)	697.17	(1011.78)

Table 26: This table shows the results for AFF with parameters (η, α) = (0.010, 0.006) for streams with approximately 5000 changepoints and different δ values, using burn-in length B=50.

Algo	Params	Values	δ	CCD	DNF	ARL1	(ARL1)	ARL0	(ARL0)
AFF	(η, α)	(0.010, 0.007)	(0.25)	0.62	0.84	55.07	(39.55)	596.72	(895.37)
AFF	(η, α)	(0.010, 0.007)	(0.50)	0.90	0.80	38.81	(31.56)	596.72	(895.37)
AFF	(η, α)	(0.010, 0.007)	(0.75)	0.95	0.75	23.11	(19.93)	596.72	(895.37)
AFF	(η, α)	(0.010, 0.007)	(1.00)	0.96	0.72	15.44	(14.18)	596.72	(895.37)
AFF	(η, α)	(0.010, 0.007)	(1.50)	0.97	0.71	9.90	(12.11)	596.72	(895.37)
AFF	(η, α)	(0.010, 0.007)	(2.00)	0.97	0.71	7.71	(11.00)	596.72	(895.37)
AFF	(η, α)	(0.010, 0.007)	(3.00)	0.97	0.70	6.02	(10.89)	596.72	(895.37)
AFF	(η, α)	(0.010, 0.007)	(4.00)	0.96	0.70	5.40	(11.60)	596.72	(895.37)

Table 27: This table shows the results for AFF with parameters (η, α) = (0.010, 0.007) for streams with approximately 5000 changepoints and different δ values, using burn-in length B=50.

Algo	Params	Values	δ	CCD	DNF	ARL1	(ARL1)	ARL0	(ARL0)
AFF	(η, α)	(0.010, 0.008)	(0.25)	0.64	0.83	54.00	(39.19)	594.55	(893.43)
AFF	(η, α)	(0.010, 0.008)	(0.50)	0.90	0.78	37.88	(31.23)	594.55	(893.43)
AFF	(η, α)	(0.010, 0.008)	(0.75)	0.95	0.74	22.56	(19.74)	594.55	(893.43)
AFF	(η, α)	(0.010, 0.008)	(1.00)	0.96	0.71	15.25	(14.71)	594.55	(893.43)
AFF	(η, α)	(0.010, 0.008)	(1.50)	0.97	0.69	9.88	(12.27)	594.55	(893.43)
AFF	(η, α)	(0.010, 0.008)	(2.00)	0.97	0.69	7.78	(11.66)	594.55	(893.43)
AFF	(η, α)	(0.010, 0.008)	(3.00)	0.97	0.68	6.13	(11.58)	594.55	(893.43)
AFF	(η, α)	(0.010, 0.008)	(4.00)	0.96	0.68	5.38	(11.56)	594.55	(893.43)

Table 28: This table shows the results for AFF with parameters (η, α) = (0.010, 0.008) for streams with approximately 5000 changepoints and different δ values, using burn-in length B=50.

Algo	Params	Values	δ	CCD	DNF	ARL1	(ARL1)	ARL0	(ARL0)
AFF	(η, α)	(0.010, 0.009)	(0.25)	0.64	0.81	52.71	(39.14)	502.06	(720.98)
AFF	(η, α)	(0.010, 0.009)	(0.50)	0.91	0.77	36.87	(30.89)	502.06	(720.98)
AFF	(η, α)	(0.010, 0.009)	(0.75)	0.95	0.72	21.93	(19.28)	502.06	(720.98)
AFF	(η, α)	(0.010, 0.009)	(1.00)	0.96	0.69	14.85	(14.15)	502.06	(720.98)
AFF	(η, α)	(0.010, 0.009)	(1.50)	0.97	0.68	9.55	(11.85)	502.06	(720.98)
AFF	(η, α)	(0.010, 0.009)	(2.00)	0.97	0.68	7.85	(12.14)	502.06	(720.98)
AFF	(η, α)	(0.010, 0.009)	(3.00)	0.97	0.67	6.04	(11.42)	502.06	(720.98)
AFF	(η, α)	(0.010, 0.009)	(4.00)	0.96	0.67	5.29	(11.17)	502.06	(720.98)

Table 29: This table shows the results for AFF with parameters (η, α) = (0.010, 0.009) for streams with approximately 5000 changepoints and different δ values, using burn-in length B=50.

Algo	Params	Values	δ	CCD	DNF	ARL1	(ARL1)	ARL0	(ARL0)
AFF	(η, α)	(0.010, 0.010)	(0.25)	0.66	0.81	52.47	(38.84)	463.64	(698.64)
AFF	(η, α)	(0.010, 0.010)	(0.50)	0.91	0.76	35.83	(30.37)	463.64	(698.64)
AFF	(η, α)	(0.010, 0.010)	(0.75)	0.96	0.71	21.33	(18.83)	463.64	(698.64)
AFF	(η, α)	(0.010, 0.010)	(1.00)	0.96	0.69	14.57	(14.03)	463.64	(698.64)
AFF	(η, α)	(0.010, 0.010)	(1.50)	0.97	0.67	9.43	(11.79)	463.64	(698.64)
AFF	(η, α)	(0.010, 0.010)	(2.00)	0.97	0.66	7.59	(11.29)	463.64	(698.64)
AFF	(η, α)	(0.010, 0.010)	(3.00)	0.97	0.66	6.06	(11.74)	463.64	(698.64)
AFF	(η, α)	(0.010, 0.010)	(4.00)	0.96	0.66	5.40	(11.77)	463.64	(698.64)

Table 30: This table shows the results for AFF with parameters (η, α) = (0.010, 0.010) for streams with approximately 5000 changepoints and different δ values, using burn-in length B=50.

Algo	Params	Values	δ	CCD	DNF	ARL1	(ARL1)	ARL0	(ARL0)
AFF	(η, α)	(0.010, 0.025)	(0.25)	0.79	0.68	43.32	(36.93)	225.22	(308.82)
AFF	(η, α)	(0.010, 0.025)	(0.50)	0.94	0.64	27.35	(25.07)	225.22	(308.82)
AFF	(η, α)	(0.010, 0.025)	(0.75)	0.96	0.60	17.39	(17.42)	225.22	(308.82)
AFF	(η, α)	(0.010, 0.025)	(1.00)	0.96	0.58	12.58	(14.10)	225.22	(308.82)
AFF	(η, α)	(0.010, 0.025)	(1.50)	0.97	0.57	9.21	(13.21)	225.22	(308.82)
AFF	(η, α)	(0.010, 0.025)	(2.00)	0.97	0.56	7.42	(12.28)	225.22	(308.82)
AFF	(η, α)	(0.010, 0.025)	(3.00)	0.97	0.56	6.23	(12.20)	225.22	(308.82)
AFF	(η, α)	(0.010, 0.025)	(4.00)	0.96	0.55	6.01	(13.25)	225.22	(308.82)

Table 31: This table shows the results for AFF with parameters (η, α) = (0.010, 0.025) for streams with approximately 5000 changepoints and different δ values, using burn-in length B=50.

Algo	Params	Values	δ	CCD	DNF	ARL1	(ARL1)	ARL0	(ARL0)
AFF	(η, α)	(0.010, 0.050)	(0.25)	0.88	0.59	34.05	(33.14)	146.05	(145.92)
AFF	(η, α)	(0.010, 0.050)	(0.50)	0.96	0.54	20.96	(21.57)	146.05	(145.92)
AFF	(η, α)	(0.010, 0.050)	(0.75)	0.97	0.52	14.11	(15.72)	146.05	(145.92)
AFF	(η, α)	(0.010, 0.050)	(1.00)	0.98	0.51	11.19	(14.10)	146.05	(145.92)
AFF	(η, α)	(0.010, 0.050)	(1.50)	0.98	0.49	7.84	(11.69)	146.05	(145.92)
AFF	(η, α)	(0.010, 0.050)	(2.00)	0.98	0.49	6.59	(11.41)	146.05	(145.92)
AFF	(η, α)	(0.010, 0.050)	(3.00)	0.98	0.49	5.96	(12.39)	146.05	(145.92)
AFF	(η, α)	(0.010, 0.050)	(4.00)	0.98	0.48	5.59	(12.49)	146.05	(145.92)

Table 32: This table shows the results for AFF with parameters (η, α) = (0.010, 0.050) for streams with approximately 5000 changepoints and different δ values, using burn-in length B=50.

Algo	Params	Values	δ	CCD	DNF	ARL1	(ARL1)	ARL0	(ARL0)
AFF	(η, α)	(0.001, 0.005)	(0.25)	0.59	0.86	55.30	(39.01)	928.78	(1253.12)
AFF	(η, α)	(0.001, 0.005)	(0.50)	0.89	0.82	37.61	(30.12)	928.78	(1253.12)
AFF	(η, α)	(0.001, 0.005)	(0.75)	0.95	0.79	21.86	(17.94)	928.78	(1253.12)
AFF	(η, α)	(0.001, 0.005)	(1.00)	0.96	0.76	14.24	(12.08)	928.78	(1253.12)
AFF	(η, α)	(0.001, 0.005)	(1.50)	0.97	0.75	8.91	(10.39)	928.78	(1253.12)
AFF	(η, α)	(0.001, 0.005)	(2.00)	0.97	0.75	6.92	(9.77)	928.78	(1253.12)
AFF	(η, α)	(0.001, 0.005)	(3.00)	0.97	0.74	4.99	(8.50)	928.78	(1253.12)
AFF	(η, α)	(0.001, 0.005)	(4.00)	0.97	0.74	4.11	(7.79)	928.78	(1253.12)

Table 33: This table shows the results for AFF with parameters (η, α) = (0.001, 0.005) for streams with approximately 5000 changepoints and different δ values, using burn-in length B=50.

Algo	Params	Values	δ	CCD	DNF	ARL1	(ARL1)	ARL0	(ARL0)
AFF	(η, α)	(0.001, 0.006)	(0.25)	0.62	0.86	54.45	(38.50)	952.18	(1267.24)
AFF	(η, α)	(0.001, 0.006)	(0.50)	0.90	0.81	36.13	(29.52)	952.18	(1267.24)
AFF	(η, α)	(0.001, 0.006)	(0.75)	0.96	0.77	20.91	(17.33)	952.18	(1267.24)
AFF	(η, α)	(0.001, 0.006)	(1.00)	0.96	0.75	14.10	(12.55)	952.18	(1267.24)
AFF	(η, α)	(0.001, 0.006)	(1.50)	0.97	0.74	8.83	(10.49)	952.18	(1267.24)
AFF	(η, α)	(0.001, 0.006)	(2.00)	0.97	0.73	6.70	(9.36)	952.18	(1267.24)
AFF	(η, α)	(0.001, 0.006)	(3.00)	0.97	0.72	4.91	(8.40)	952.18	(1267.24)
AFF	(η, α)	(0.001, 0.006)	(4.00)	0.97	0.72	4.15	(8.08)	952.18	(1267.24)

Table 34: This table shows the results for AFF with parameters (η, α) = (0.001, 0.006) for streams with approximately 5000 changepoints and different δ values, using burn-in length B=50.

Algo	Params	Values	δ	CCD	DNF	ARL1	(ARL1)	ARL0	(ARL0)
AFF	(η, α)	(0.001, 0.007)	(0.25)	0.64	0.84	53.13	(38.27)	832.73	(1163.07)
AFF	(η, α)	(0.001, 0.007)	(0.50)	0.91	0.80	35.22	(29.03)	832.73	(1163.07)
AFF	(η, α)	(0.001, 0.007)	(0.75)	0.96	0.76	20.36	(16.95)	832.73	(1163.07)
AFF	(η, α)	(0.001, 0.007)	(1.00)	0.96	0.74	13.75	(12.42)	832.73	(1163.07)
AFF	(η, α)	(0.001, 0.007)	(1.50)	0.97	0.72	8.67	(10.46)	832.73	(1163.07)
AFF	(η, α)	(0.001, 0.007)	(2.00)	0.97	0.71	6.50	(8.72)	832.73	(1163.07)
AFF	(η, α)	(0.001, 0.007)	(3.00)	0.97	0.71	4.84	(8.25)	832.73	(1163.07)
AFF	(η, α)	(0.001, 0.007)	(4.00)	0.97	0.71	4.12	(8.29)	832.73	(1163.07)

Table 35: This table shows the results for AFF with parameters (η, α) = (0.001, 0.007) for streams with approximately 5000 changepoints and different δ values, using burn-in length B=50.

Algo	Params	Values	δ	CCD	DNF	ARL1	(ARL1)	ARL0	(ARL0)
AFF	(η, α)	(0.001, 0.008)	(0.25)	0.64	0.84	53.22	(38.16)	793.05	(1095.04)
AFF	(η, α)	(0.001, 0.008)	(0.50)	0.91	0.78	34.69	(28.61)	793.05	(1095.04)
AFF	(η, α)	(0.001, 0.008)	(0.75)	0.95	0.74	19.90	(16.78)	793.05	(1095.04)
AFF	(η, α)	(0.001, 0.008)	(1.00)	0.96	0.72	13.67	(12.95)	793.05	(1095.04)
AFF	(η, α)	(0.001, 0.008)	(1.50)	0.97	0.71	8.65	(10.54)	793.05	(1095.04)
AFF	(η, α)	(0.001, 0.008)	(2.00)	0.97	0.71	6.57	(9.28)	793.05	(1095.04)
AFF	(η, α)	(0.001, 0.008)	(3.00)	0.97	0.70	4.83	(8.35)	793.05	(1095.04)
AFF	(η, α)	(0.001, 0.008)	(4.00)	0.97	0.69	4.09	(8.12)	793.05	(1095.04)

Table 36: This table shows the results for AFF with parameters (η, α) = (0.001, 0.008) for streams with approximately 5000 changepoints and different δ values, using burn-in length B=50.

Algo	Params	Values	δ	CCD	DNF	ARL1	(ARL1)	ARL0	(ARL0)
AFF	(η, α)	(0.001, 0.009)	(0.25)	0.66	0.82	51.75	(37.76)	782.56	(1051.46)
AFF	(η, α)	(0.001, 0.009)	(0.50)	0.91	0.77	33.56	(27.71)	782.56	(1051.46)
AFF	(η, α)	(0.001, 0.009)	(0.75)	0.95	0.73	19.57	(16.94)	782.56	(1051.46)
AFF	(η, α)	(0.001, 0.009)	(1.00)	0.96	0.72	13.50	(12.91)	782.56	(1051.46)
AFF	(η, α)	(0.001, 0.009)	(1.50)	0.97	0.70	8.39	(10.15)	782.56	(1051.46)
AFF	(η, α)	(0.001, 0.009)	(2.00)	0.97	0.70	6.49	(9.19)	782.56	(1051.46)
AFF	(η, α)	(0.001, 0.009)	(3.00)	0.97	0.69	4.90	(8.93)	782.56	(1051.46)
AFF	(η, α)	(0.001, 0.009)	(4.00)	0.97	0.68	4.16	(8.74)	782.56	(1051.46)

Table 37: This table shows the results for AFF with parameters (η, α) = (0.001, 0.009) for streams with approximately 5000 changepoints and different δ values, using burn-in length B=50.

Algo	Params	Values	δ	CCD	DNF	ARL1	(ARL1)	ARL0	(ARL0)
AFF	(η, α)	(0.001, 0.010)	(0.25)	0.66	0.82	51.06	(37.70)	729.64	(1004.34)
AFF	(η, α)	(0.001, 0.010)	(0.50)	0.91	0.76	33.35	(28.01)	729.64	(1004.34)
AFF	(η, α)	(0.001, 0.010)	(0.75)	0.95	0.72	19.50	(17.11)	729.64	(1004.34)
AFF	(η, α)	(0.001, 0.010)	(1.00)	0.97	0.71	13.23	(12.84)	729.64	(1004.34)
AFF	(η, α)	(0.001, 0.010)	(1.50)	0.97	0.69	8.37	(10.28)	729.64	(1004.34)
AFF	(η, α)	(0.001, 0.010)	(2.00)	0.97	0.68	6.55	(9.73)	729.64	(1004.34)
AFF	(η, α)	(0.001, 0.010)	(3.00)	0.97	0.68	4.77	(8.60)	729.64	(1004.34)
AFF	(η, α)	(0.001, 0.010)	(4.00)	0.97	0.67	4.19	(8.68)	729.64	(1004.34)

Table 38: This table shows the results for AFF with parameters (η, α) = (0.001, 0.010) for streams with approximately 5000 changepoints and different δ values, using burn-in length B=50.

Algo	Params	Values	δ	CCD	DNF	ARL1	(ARL1)	ARL0	(ARL0)
AFF	(η, α)	(0.001, 0.025)	(0.25)	0.78	0.72	44.72	(36.12)	416.94	(546.95)
AFF	(η, α)	(0.001, 0.025)	(0.50)	0.94	0.66	27.68	(25.00)	416.94	(546.95)
AFF	(η, α)	(0.001, 0.025)	(0.75)	0.96	0.63	16.45	(15.78)	416.94	(546.95)
AFF	(η, α)	(0.001, 0.025)	(1.00)	0.97	0.61	12.01	(13.58)	416.94	(546.95)
AFF	(η, α)	(0.001, 0.025)	(1.50)	0.97	0.59	7.94	(11.72)	416.94	(546.95)
AFF	(η, α)	(0.001, 0.025)	(2.00)	0.98	0.59	6.27	(10.87)	416.94	(546.95)
AFF	(η, α)	(0.001, 0.025)	(3.00)	0.98	0.58	4.90	(10.19)	416.94	(546.95)
AFF	(η, α)	(0.001, 0.025)	(4.00)	0.97	0.57	4.34	(9.99)	416.94	(546.95)

Table 39: This table shows the results for AFF with parameters (η, α) = (0.001, 0.025) for streams with approximately 5000 changepoints and different δ values, using burn-in length B=50.

Algo	Params	Values	δ	CCD	DNF	ARL1	(ARL1)	ARL0	(ARL0)
AFF	(η, α)	(0.001, 0.050)	(0.25)	0.86	0.64	38.09	(33.01)	301.30	(379.49)
AFF	(η, α)	(0.001, 0.050)	(0.50)	0.96	0.58	22.55	(21.89)	301.30	(379.49)
AFF	(η, α)	(0.001, 0.050)	(0.75)	0.97	0.55	14.59	(15.96)	301.30	(379.49)
AFF	(η, α)	(0.001, 0.050)	(1.00)	0.98	0.54	11.12	(14.32)	301.30	(379.49)
AFF	(η, α)	(0.001, 0.050)	(1.50)	0.98	0.52	7.51	(12.08)	301.30	(379.49)
AFF	(η, α)	(0.001, 0.050)	(2.00)	0.98	0.52	5.72	(10.40)	301.30	(379.49)
AFF	(η, α)	(0.001, 0.050)	(3.00)	0.98	0.51	4.45	(9.64)	301.30	(379.49)
AFF	(η, α)	(0.001, 0.050)	(4.00)	0.98	0.50	4.30	(10.39)	301.30	(379.49)

Table 40: This table shows the results for AFF with parameters (η, α) = (0.001, 0.050) for streams with approximately 5000 changepoints and different δ values, using burn-in length B=50.

3.1.1 Discussion

These tables provide results for how each algorithm, for a selection of control parameters, performs (in terms of ARL1) for changes of different size $\delta \in \{0.25, 0.50, 0.75, 1.00, 1.50, 2.00, 3.00, 4.00\}$. Recall that the ARL0 values are all the same (for a particular choice of control parameters), since ARL0 is computed on streams not containing any changepoints.

3.2 Tables of each algorithm for a fixed δ, as parameter pair varies

In this section, each table provides the results for a particular algorithm, for a given stream containing changepoints of a fixed magnitude, but the parameters of the algorithm are varied in each table.

Algo	Params	Values	CCD	DNF	ARL1	(ARL1)	ARL0	(ARL0)
CUSUM	(k, h)	(0.25, 8.01)	0.72	0.84	50.53	(36.95)	281.57	(431.99)
CUSUM	(k, h)	(0.50, 4.77)	0.60	0.84	52.96	(39.19)	357.41	(582.96)
CUSUM	(k, h)	(0.75, 3.34)	0.51	0.85	54.60	(40.52)	440.62	(757.59)
CUSUM	(k, h)	(1.00, 2.52)	0.44	0.86	57.40	(41.14)	479.24	(848.11)
CUSUM	(k, h)	(1.25, 1.99)	0.39	0.86	58.89	(41.88)	526.88	(906.81)
CUSUM	(k, h)	(1.50, 1.61)	0.36	0.86	60.24	(41.82)	493.38	(855.65)
EWMA	(r, L)	(1.00, 3.090)	0.29	0.90	64.82	(43.04)	661.37	(987.72)
EWMA	(r, L)	(0.75, 3.087)	0.33	0.88	62.23	(41.73)	676.26	(1072.18)
EWMA	(r, L)	(0.50, 3.071)	0.39	0.88	60.23	(41.54)	638.20	(983.45)
EWMA	(r, L)	(0.40, 3.054)	0.44	0.87	57.75	(41.60)	564.67	(886.39)
EWMA	(r, L)	(0.30, 3.023)	0.47	0.87	57.02	(41.50)	506.76	(819.15)
EWMA	(r, L)	(0.25, 2.998)	0.51	0.87	55.79	(41.10)	502.91	(821.13)
EWMA	(r, L)	(0.20, 2.962)	0.55	0.86	54.43	(40.17)	493.38	(799.77)
EWMA	(r, L)	(0.10, 2.814)	0.64	0.86	52.87	(38.97)	407.02	(690.85)
EWMA	(r, L)	(0.05, 2.615)	0.87	0.41	17.23	(32.28)	356.06	(568.99)
EWMA	(r, L)	(0.03, 2.437)	0.99	0.35	2.89	(12.01)	312.45	(456.83)
AFF	(η, α)	(0.100, 0.005)	0.55	0.88	56.67	(39.80)	670.03	(972.68)
AFF	(η, α)	(0.100, 0.006)	0.58	0.86	55.24	(39.68)	576.34	(893.73)
AFF	(η, α)	(0.100, 0.007)	0.60	0.85	54.23	(39.64)	494.74	(765.15)
AFF	(η, α)	(0.100, 0.008)	0.62	0.83	53.49	(39.69)	396.80	(629.73)
AFF	(η, α)	(0.100, 0.009)	0.64	0.82	52.88	(39.13)	388.93	(634.65)
AFF	(η, α)	(0.100, 0.010)	0.65	0.81	51.83	(38.99)	352.77	(557.18)
AFF	(η, α)	(0.100, 0.025)	0.79	0.70	44.19	(36.70)	149.56	(199.33)
AFF	(η, α)	(0.100, 0.050)	0.90	0.60	33.91	(32.19)	87.67	(95.61)
AFF	(η, α)	(0.010, 0.005)	0.58	0.86	56.65	(40.34)	766.12	(1051.05)
AFF	(η, α)	(0.010, 0.006)	0.60	0.85	56.04	(40.08)	697.17	(1011.78)
AFF	(η, α)	(0.010, 0.007)	0.62	0.84	55.07	(39.55)	596.72	(895.37)
AFF	(η, α)	(0.010, 0.008)	0.64	0.83	54.00	(39.19)	594.55	(893.43)
AFF	(η, α)	(0.010, 0.009)	0.64	0.81	52.71	(39.14)	502.06	(720.98)
AFF	(η, α)	(0.010, 0.010)	0.66	0.81	52.47	(38.84)	463.64	(698.64)
AFF	(η, α)	(0.010, 0.025)	0.79	0.68	43.32	(36.93)	225.22	(308.82)
AFF	(η, α)	(0.010, 0.050)	0.88	0.59	34.05	(33.14)	146.05	(145.92)
AFF	(η, α)	(0.001, 0.005)	0.59	0.86	55.30	(39.01)	928.78	(1253.12)
AFF	(η, α)	(0.001, 0.006)	0.62	0.86	54.45	(38.50)	952.18	(1267.24)
AFF	(η, α)	(0.001, 0.007)	0.64	0.84	53.13	(38.27)	832.73	(1163.07)
AFF	(η, α)	(0.001, 0.008)	0.64	0.84	53.22	(38.16)	793.05	(1095.04)
AFF	(η, α)	(0.001, 0.009)	0.66	0.82	51.75	(37.76)	782.56	(1051.46)
AFF	(η, α)	(0.001, 0.010)	0.66	0.82	51.06	(37.70)	729.64	(1004.34)
AFF	(η, α)	(0.001, 0.025)	0.78	0.72	44.72	(36.12)	416.94	(546.95)
AFF	(η, α)	(0.001, 0.050)	0.86	0.64	38.09	(33.01)	301.30	(379.49)

Table 41: Summary of detection efficiency for algorithms listed, over 750,000 observations with approximately 5000 changepoints, with $\delta = 0.25$ and B=50.

Algo	Params	Values	CCD	DNF	ARL1	(ARL1)	ARL0	(ARL0)
CUSUM	(k, h)	(0.25, 8.01)	0.91	0.79	32.11	(26.32)	281.57	(431.99)
CUSUM	(k, h)	(0.50, 4.77)	0.84	0.79	36.32	(31.83)	357.41	(582.96)
CUSUM	(k, h)	(0.75, 3.34)	0.75	0.82	40.73	(35.43)	440.62	(757.59)
CUSUM	(k, h)	(1.00, 2.52)	0.67	0.83	45.18	(37.59)	479.24	(848.11)
CUSUM	(k, h)	(1.25, 1.99)	0.61	0.84	48.16	(39.13)	526.88	(906.81)
CUSUM	(k, h)	(1.50, 1.61)	0.56	0.85	51.10	(39.99)	493.38	(855.65)
EWMA	(r, L)	(1.00, 3.090)	0.44	0.88	56.59	(41.03)	661.37	(987.72)
EWMA	(r, L)	(0.75, 3.087)	0.51	0.87	52.72	(40.34)	676.26	(1072.18)
EWMA	(r, L)	(0.50, 3.071)	0.62	0.85	47.67	(38.90)	638.20	(983.45)
EWMA	(r, L)	(0.40, 3.054)	0.67	0.85	45.55	(38.32)	564.67	(886.39)
EWMA	(r, L)	(0.30, 3.023)	0.74	0.84	42.11	(36.19)	506.76	(819.15)
EWMA	(r, L)	(0.25, 2.998)	0.77	0.83	40.81	(35.31)	502.91	(821.13)
EWMA	(r, L)	(0.20, 2.962)	0.81	0.83	38.43	(33.22)	493.38	(799.77)
EWMA	(r, L)	(0.10, 2.814)	0.89	0.81	34.29	(29.47)	407.02	(690.85)
EWMA	(r, L)	(0.05, 2.615)	0.98	0.36	5.82	(16.41)	356.06	(568.99)
EWMA	(r, L)	(0.03, 2.437)	1.00	0.34	1.91	(7.29)	312.45	(456.83)
AFF	(η, α)	(0.100, 0.005)	0.86	0.84	40.61	(32.27)	670.03	(972.68)
AFF	(η, α)	(0.100, 0.006)	0.86	0.83	39.28	(31.83)	576.34	(893.73)
AFF	(η, α)	(0.100, 0.007)	0.88	0.81	38.08	(31.13)	494.74	(765.15)
AFF	(η, α)	(0.100, 0.008)	0.89	0.79	37.21	(30.90)	396.80	(629.73)
AFF	(η, α)	(0.100, 0.009)	0.89	0.78	36.22	(30.23)	388.93	(634.65)
AFF	(η, α)	(0.100, 0.010)	0.90	0.76	35.16	(29.78)	352.77	(557.18)
AFF	(η, α)	(0.100, 0.025)	0.93	0.65	28.11	(26.19)	149.56	(199.33)
AFF	(η, α)	(0.100, 0.050)	0.96	0.56	22.16	(23.01)	87.67	(95.61)
AFF	(η, α)	(0.010, 0.005)	0.88	0.83	42.02	(32.92)	766.12	(1051.05)
AFF	(η, α)	(0.010, 0.006)	0.89	0.82	40.08	(32.36)	697.17	(1011.78)
AFF	(η, α)	(0.010, 0.007)	0.90	0.80	38.81	(31.56)	596.72	(895.37)
AFF	(η, α)	(0.010, 0.008)	0.90	0.78	37.88	(31.23)	594.55	(893.43)
AFF	(η, α)	(0.010, 0.009)	0.91	0.77	36.87	(30.89)	502.06	(720.98)
AFF	(η, α)	(0.010, 0.010)	0.91	0.76	35.83	(30.37)	463.64	(698.64)
AFF	(η, α)	(0.010, 0.025)	0.94	0.64	27.35	(25.07)	225.22	(308.82)
AFF	(η, α)	(0.010, 0.050)	0.96	0.54	20.96	(21.57)	146.05	(145.92)
AFF	(η, α)	(0.001, 0.005)	0.89	0.82	37.61	(30.12)	928.78	(1253.12)
AFF	(η, α)	(0.001, 0.006)	0.90	0.81	36.13	(29.52)	952.18	(1267.24)
AFF	(η, α)	(0.001, 0.007)	0.91	0.80	35.22	(29.03)	832.73	(1163.07)
AFF	(η, α)	(0.001, 0.008)	0.91	0.78	34.69	(28.61)	793.05	(1095.04)
AFF	(η, α)	(0.001, 0.009)	0.91	0.77	33.56	(27.71)	782.56	(1051.46)
AFF	(η, α)	(0.001, 0.010)	0.91	0.76	33.35	(28.01)	729.64	(1004.34)
AFF	(η, α)	(0.001, 0.025)	0.94	0.66	27.68	(25.00)	416.94	(546.95)
AFF	(η, α)	(0.001, 0.050)	0.96	0.58	22.55	(21.89)	301.30	(379.49)

Table 42: Summary of detection efficiency for algorithms listed, over 750,000 observations with approximately 5000 changepoints, with $\delta = 0.5$ and B=50.

Algo	Params	Values	CCD	DNF	ARL1	(ARL1)	ARL0	(ARL0)
CUSUM	(k, h)	(0.25, 8.01)	0.95	0.75	20.06	(16.60)	281.57	(431.99)
CUSUM	(k, h)	(0.50, 4.77)	0.94	0.77	21.83	(21.32)	357.41	(582.96)
CUSUM	(k, h)	(0.75, 3.34)	0.91	0.79	26.81	(27.03)	440.62	(757.59)
CUSUM	(k, h)	(1.00, 2.52)	0.84	0.80	31.72	(30.80)	479.24	(848.11)
CUSUM	(k, h)	(1.25, 1.99)	0.78	0.81	36.58	(34.19)	526.88	(906.81)
CUSUM	(k, h)	(1.50, 1.61)	0.71	0.82	40.66	(36.66)	493.38	(855.65)
EWMA	(r, L)	(1.00, 3.090)	0.56	0.87	48.72	(39.34)	661.37	(987.72)
EWMA	(r, L)	(0.75, 3.087)	0.67	0.85	42.66	(37.53)	676.26	(1072.18)
EWMA	(r, L)	(0.50, 3.071)	0.80	0.84	36.97	(34.81)	638.20	(983.45)
EWMA	(r, L)	(0.40, 3.054)	0.85	0.82	32.87	(31.69)	564.67	(886.39)
EWMA	(r, L)	(0.30, 3.023)	0.90	0.81	28.43	(28.06)	506.76	(819.15)
EWMA	(r, L)	(0.25, 2.998)	0.91	0.80	26.26	(26.29)	502.91	(821.13)
EWMA	(r, L)	(0.20, 2.962)	0.93	0.80	23.96	(23.71)	493.38	(799.77)
EWMA	(r, L)	(0.10, 2.814)	0.95	0.77	19.82	(17.75)	407.02	(690.85)
EWMA	(r, L)	(0.05, 2.615)	0.99	0.35	2.71	(7.97)	356.06	(568.99)
EWMA	(r, L)	(0.03, 2.437)	1.00	0.34	1.39	(3.39)	312.45	(456.83)
AFF	(η, α)	(0.100, 0.005)	0.96	0.81	24.13	(20.46)	670.03	(972.68)
AFF	(η, α)	(0.100, 0.006)	0.96	0.79	23.12	(19.85)	576.34	(893.73)
AFF	(η, α)	(0.100, 0.007)	0.96	0.78	22.42	(19.41)	494.74	(765.15)
AFF	(η, α)	(0.100, 0.008)	0.96	0.76	21.86	(19.18)	396.80	(629.73)
AFF	(η, α)	(0.100, 0.009)	0.96	0.75	21.28	(18.78)	388.93	(634.65)
AFF	(η, α)	(0.100, 0.010)	0.96	0.74	20.92	(18.71)	352.77	(557.18)
AFF	(η, α)	(0.100, 0.025)	0.96	0.62	17.04	(17.22)	149.56	(199.33)
AFF	(η, α)	(0.100, 0.050)	0.97	0.54	14.96	(17.53)	87.67	(95.61)
AFF	(η, α)	(0.010, 0.005)	0.96	0.79	25.03	(21.12)	766.12	(1051.05)
AFF	(η, α)	(0.010, 0.006)	0.95	0.76	23.68	(19.94)	697.17	(1011.78)
AFF	(η, α)	(0.010, 0.007)	0.95	0.75	23.11	(19.93)	596.72	(895.37)
AFF	(η, α)	(0.010, 0.008)	0.95	0.74	22.56	(19.74)	594.55	(893.43)
AFF	(η, α)	(0.010, 0.009)	0.95	0.72	21.93	(19.28)	502.06	(720.98)
AFF	(η, α)	(0.010, 0.010)	0.96	0.71	21.33	(18.83)	463.64	(698.64)
AFF	(η, α)	(0.010, 0.025)	0.96	0.60	17.39	(17.42)	225.22	(308.82)
AFF	(η, α)	(0.010, 0.050)	0.97	0.52	14.11	(15.72)	146.05	(145.92)
AFF	(η, α)	(0.001, 0.005)	0.95	0.79	21.86	(17.94)	928.78	(1253.12)
AFF	(η, α)	(0.001, 0.006)	0.96	0.77	20.91	(17.33)	952.18	(1267.24)
AFF	(η, α)	(0.001, 0.007)	0.96	0.76	20.36	(16.95)	832.73	(1163.07)
AFF	(η, α)	(0.001, 0.008)	0.95	0.74	19.90	(16.78)	793.05	(1095.04)
AFF	(η, α)	(0.001, 0.009)	0.95	0.73	19.57	(16.94)	782.56	(1051.46)
AFF	(η, α)	(0.001, 0.010)	0.95	0.72	19.50	(17.11)	729.64	(1004.34)
AFF	(η, α)	(0.001, 0.025)	0.96	0.63	16.45	(15.78)	416.94	(546.95)
AFF	(η, α)	(0.001, 0.050)	0.97	0.55	14.59	(15.96)	301.30	(379.49)

Table 43: Summary of detection efficiency for algorithms listed, over 750,000 observations with approximately 5000 changepoints, with $\delta = 0.75$ and B=50.

Algo	Params	Values	CCD	DNF	ARL1	(ARL1)	ARL0	(ARL0)
CUSUM	(k, h)	(0.25, 8.01)	0.96	0.74	14.62	(13.51)	281.57	(431.99)
CUSUM	(k, h)	(0.50, 4.77)	0.95	0.75	14.38	(15.07)	357.41	(582.96)
CUSUM	(k, h)	(0.75, 3.34)	0.94	0.76	16.44	(18.06)	440.62	(757.59)
CUSUM	(k, h)	(1.00, 2.52)	0.92	0.78	20.20	(22.32)	479.24	(848.11)
CUSUM	(k, h)	(1.25, 1.99)	0.88	0.79	25.58	(27.33)	526.88	(906.81)
CUSUM	(k, h)	(1.50, 1.61)	0.83	0.80	30.37	(31.35)	493.38	(855.65)
EWMA	(r, L)	(1.00, 3.090)	0.68	0.86	42.14	(37.63)	661.37	(987.72)
EWMA	(r, L)	(0.75, 3.087)	0.80	0.84	33.73	(33.26)	676.26	(1072.18)
EWMA	(r, L)	(0.50, 3.071)	0.90	0.81	24.34	(26.18)	638.20	(983.45)
EWMA	(r, L)	(0.40, 3.054)	0.92	0.80	20.13	(21.81)	564.67	(886.39)
EWMA	(r, L)	(0.30, 3.023)	0.94	0.79	17.26	(18.71)	506.76	(819.15)
EWMA	(r, L)	(0.25, 2.998)	0.94	0.79	16.14	(17.54)	502.91	(821.13)
EWMA	(r, L)	(0.20, 2.962)	0.95	0.78	14.73	(15.44)	493.38	(799.77)
EWMA	(r, L)	(0.10, 2.814)	0.96	0.75	13.06	(11.75)	407.02	(690.85)
EWMA	(r, L)	(0.05, 2.615)	1.00	0.35	2.06	(5.89)	356.06	(568.99)
EWMA	(r, L)	(0.03, 2.437)	1.00	0.34	1.30	(3.20)	312.45	(456.83)
AFF	(η, α)	(0.100, 0.005)	0.97	0.80	15.24	(13.01)	670.03	(972.68)
AFF	(η, α)	(0.100, 0.006)	0.97	0.78	14.89	(13.02)	576.34	(893.73)
AFF	(η, α)	(0.100, 0.007)	0.97	0.76	14.41	(13.04)	494.74	(765.15)
AFF	(η, α)	(0.100, 0.008)	0.96	0.75	14.10	(12.86)	396.80	(629.73)
AFF	(η, α)	(0.100, 0.009)	0.97	0.73	13.94	(13.08)	388.93	(634.65)
AFF	(η, α)	(0.100, 0.010)	0.96	0.72	13.79	(13.35)	352.77	(557.18)
AFF	(η, α)	(0.100, 0.025)	0.97	0.61	12.37	(14.35)	149.56	(199.33)
AFF	(η, α)	(0.100, 0.050)	0.98	0.53	11.31	(15.09)	87.67	(95.61)
AFF	(η, α)	(0.010, 0.005)	0.97	0.77	16.47	(14.63)	766.12	(1051.05)
AFF	(η, α)	(0.010, 0.006)	0.96	0.74	15.95	(14.66)	697.17	(1011.78)
AFF	(η, α)	(0.010, 0.007)	0.96	0.72	15.44	(14.18)	596.72	(895.37)
AFF	(η, α)	(0.010, 0.008)	0.96	0.71	15.25	(14.71)	594.55	(893.43)
AFF	(η, α)	(0.010, 0.009)	0.96	0.69	14.85	(14.15)	502.06	(720.98)
AFF	(η, α)	(0.010, 0.010)	0.96	0.69	14.57	(14.03)	463.64	(698.64)
AFF	(η, α)	(0.010, 0.025)	0.96	0.58	12.58	(14.10)	225.22	(308.82)
AFF	(η, α)	(0.010, 0.050)	0.98	0.51	11.19	(14.10)	146.05	(145.92)
AFF	(η, α)	(0.001, 0.005)	0.96	0.76	14.24	(12.08)	928.78	(1253.12)
AFF	(η, α)	(0.001, 0.006)	0.96	0.75	14.10	(12.55)	952.18	(1267.24)
AFF	(η, α)	(0.001, 0.007)	0.96	0.74	13.75	(12.42)	832.73	(1163.07)
AFF	(η, α)	(0.001, 0.008)	0.96	0.72	13.67	(12.95)	793.05	(1095.04)
AFF	(η, α)	(0.001, 0.009)	0.96	0.72	13.50	(12.91)	782.56	(1051.46)
AFF	(η, α)	(0.001, 0.010)	0.97	0.71	13.23	(12.84)	729.64	(1004.34)
AFF	(η, α)	(0.001, 0.025)	0.97	0.61	12.01	(13.58)	416.94	(546.95)
AFF	(η, α)	(0.001, 0.050)	0.98	0.54	11.12	(14.32)	301.30	(379.49)

Table 44: Summary of detection efficiency for algorithms listed, over 750,000 observations with approximately 5000 changepoints, with $\delta = 1$ and B=50.

Algo	Params	Values	CCD	DNF	ARL1	(ARL1)	ARL0	(ARL0)
CUSUM	(k, h)	(0.25, 8.01)	0.97	0.74	10.33	(11.20)	281.57	(431.99)
CUSUM	(k, h)	(0.50, 4.77)	0.96	0.74	8.97	(11.57)	357.41	(582.96)
CUSUM	(k, h)	(0.75, 3.34)	0.95	0.75	8.92	(12.33)	440.62	(757.59)
CUSUM	(k, h)	(1.00, 2.52)	0.93	0.76	10.04	(14.48)	479.24	(848.11)
CUSUM	(k, h)	(1.25, 1.99)	0.91	0.76	11.50	(15.92)	526.88	(906.81)
CUSUM	(k, h)	(1.50, 1.61)	0.90	0.77	13.97	(18.66)	493.38	(855.65)
EWMA	(r, L)	(1.00, 3.090)	0.85	0.83	25.31	(28.20)	661.37	(987.72)
EWMA	(r, L)	(0.75, 3.087)	0.91	0.80	16.12	(19.92)	676.26	(1072.18)
EWMA	(r, L)	(0.50, 3.071)	0.93	0.79	10.89	(14.71)	638.20	(983.45)
EWMA	(r, L)	(0.40, 3.054)	0.94	0.79	9.82	(13.44)	564.67	(886.39)
EWMA	(r, L)	(0.30, 3.023)	0.95	0.78	9.09	(12.69)	506.76	(819.15)
EWMA	(r, L)	(0.25, 2.998)	0.96	0.78	8.42	(11.00)	502.91	(821.13)
EWMA	(r, L)	(0.20, 2.962)	0.96	0.77	8.03	(9.72)	493.38	(799.77)
EWMA	(r, L)	(0.10, 2.814)	0.97	0.71	8.49	(10.54)	407.02	(690.85)
EWMA	(r, L)	(0.05, 2.615)	1.00	0.35	1.59	(4.10)	356.06	(568.99)
EWMA	(r, L)	(0.03, 2.437)	1.00	0.34	1.19	(3.02)	312.45	(456.83)
AFF	(η, α)	(0.100, 0.005)	0.97	0.78	9.07	(9.76)	670.03	(972.68)
AFF	(η, α)	(0.100, 0.006)	0.97	0.76	9.00	(10.08)	576.34	(893.73)
AFF	(η, α)	(0.100, 0.007)	0.97	0.75	8.84	(10.12)	494.74	(765.15)
AFF	(η, α)	(0.100, 0.008)	0.97	0.73	8.77	(10.32)	396.80	(629.73)
AFF	(η, α)	(0.100, 0.009)	0.97	0.72	8.74	(10.65)	388.93	(634.65)
AFF	(η, α)	(0.100, 0.010)	0.97	0.71	8.78	(10.84)	352.77	(557.18)
AFF	(η, α)	(0.100, 0.025)	0.98	0.60	8.47	(12.14)	149.56	(199.33)
AFF	(η, α)	(0.100, 0.050)	0.98	0.52	8.19	(13.03)	87.67	(95.61)
AFF	(η, α)	(0.010, 0.005)	0.97	0.74	10.00	(11.35)	766.12	(1051.05)
AFF	(η, α)	(0.010, 0.006)	0.97	0.72	9.73	(11.03)	697.17	(1011.78)
AFF	(η, α)	(0.010, 0.007)	0.97	0.71	9.90	(12.11)	596.72	(895.37)
AFF	(η, α)	(0.010, 0.008)	0.97	0.69	9.88	(12.27)	594.55	(893.43)
AFF	(η, α)	(0.010, 0.009)	0.97	0.68	9.55	(11.85)	502.06	(720.98)
AFF	(η, α)	(0.010, 0.010)	0.97	0.67	9.43	(11.79)	463.64	(698.64)
AFF	(η, α)	(0.010, 0.025)	0.97	0.57	9.21	(13.21)	225.22	(308.82)
AFF	(η, α)	(0.010, 0.050)	0.98	0.49	7.84	(11.69)	146.05	(145.92)
AFF	(η, α)	(0.001, 0.005)	0.97	0.75	8.91	(10.39)	928.78	(1253.12)
AFF	(η, α)	(0.001, 0.006)	0.97	0.74	8.83	(10.49)	952.18	(1267.24)
AFF	(η, α)	(0.001, 0.007)	0.97	0.72	8.67	(10.46)	832.73	(1163.07)
AFF	(η, α)	(0.001, 0.008)	0.97	0.71	8.65	(10.54)	793.05	(1095.04)
AFF	(η, α)	(0.001, 0.009)	0.97	0.70	8.39	(10.15)	782.56	(1051.46)
AFF	(η, α)	(0.001, 0.010)	0.97	0.69	8.37	(10.28)	729.64	(1004.34)
AFF	(η, α)	(0.001, 0.025)	0.97	0.59	7.94	(11.72)	416.94	(546.95)
AFF	(η, α)	(0.001, 0.050)	0.98	0.52	7.51	(12.08)	301.30	(379.49)

Table 45: Summary of detection efficiency for algorithms listed, over 750,000 observations with approximately 5000 changepoints, with $\delta = 1.5$ and B=50.

Algo	Params	Values	CCD	DNF	ARL1	(ARL1)	ARL0	(ARL0)
CUSUM	(k, h)	(0.25, 8.01)	0.98	0.73	8.24	(9.99)	281.57	(431.99)
CUSUM	(k, h)	(0.50, 4.77)	0.97	0.74	7.05	(10.74)	357.41	(582.96)
CUSUM	(k, h)	(0.75, 3.34)	0.95	0.75	6.61	(10.80)	440.62	(757.59)
CUSUM	(k, h)	(1.00, 2.52)	0.93	0.75	6.74	(11.86)	479.24	(848.11)
CUSUM	(k, h)	(1.25, 1.99)	0.91	0.75	7.19	(12.90)	526.88	(906.81)
CUSUM	(k, h)	(1.50, 1.61)	0.89	0.76	8.03	(14.35)	493.38	(855.65)
EWMA	(r, L)	(1.00, 3.090)	0.87	0.80	12.79	(18.06)	661.37	(987.72)
EWMA	(r, L)	(0.75, 3.087)	0.90	0.79	8.59	(13.56)	676.26	(1072.18)
EWMA	(r, L)	(0.50, 3.071)	0.92	0.78	7.08	(12.15)	638.20	(983.45)
EWMA	(r, L)	(0.40, 3.054)	0.94	0.78	6.82	(11.62)	564.67	(886.39)
EWMA	(r, L)	(0.30, 3.023)	0.95	0.78	6.49	(10.87)	506.76	(819.15)
EWMA	(r, L)	(0.25, 2.998)	0.96	0.77	6.17	(9.58)	502.91	(821.13)
EWMA	(r, L)	(0.20, 2.962)	0.96	0.77	6.03	(8.82)	493.38	(799.77)
EWMA	(r, L)	(0.10, 2.814)	0.97	0.66	6.42	(9.25)	407.02	(690.85)
EWMA	(r, L)	(0.05, 2.615)	1.00	0.35	1.41	(3.70)	356.06	(568.99)
EWMA	(r, L)	(0.03, 2.437)	1.00	0.34	1.14	(2.22)	312.45	(456.83)
AFF	(η, α)	(0.100, 0.005)	0.98	0.78	7.06	(9.48)	670.03	(972.68)
AFF	(η, α)	(0.100, 0.006)	0.98	0.76	6.98	(9.33)	576.34	(893.73)
AFF	(η, α)	(0.100, 0.007)	0.98	0.74	7.09	(9.85)	494.74	(765.15)
AFF	(η, α)	(0.100, 0.008)	0.97	0.73	7.03	(10.02)	396.80	(629.73)
AFF	(η, α)	(0.100, 0.009)	0.97	0.72	7.02	(10.11)	388.93	(634.65)
AFF	(η, α)	(0.100, 0.010)	0.97	0.70	7.00	(9.97)	352.77	(557.18)
AFF	(η, α)	(0.100, 0.025)	0.97	0.60	6.88	(10.88)	149.56	(199.33)
AFF	(η, α)	(0.100, 0.050)	0.98	0.52	6.58	(11.34)	87.67	(95.61)
AFF	(η, α)	(0.010, 0.005)	0.97	0.74	7.58	(9.99)	766.12	(1051.05)
AFF	(η, α)	(0.010, 0.006)	0.97	0.72	7.85	(11.30)	697.17	(1011.78)
AFF	(η, α)	(0.010, 0.007)	0.97	0.71	7.71	(11.00)	596.72	(895.37)
AFF	(η, α)	(0.010, 0.008)	0.97	0.69	7.78	(11.66)	594.55	(893.43)
AFF	(η, α)	(0.010, 0.009)	0.97	0.68	7.85	(12.14)	502.06	(720.98)
AFF	(η, α)	(0.010, 0.010)	0.97	0.66	7.59	(11.29)	463.64	(698.64)
AFF	(η, α)	(0.010, 0.025)	0.97	0.56	7.42	(12.28)	225.22	(308.82)
AFF	(η, α)	(0.010, 0.050)	0.98	0.49	6.59	(11.41)	146.05	(145.92)
AFF	(η, α)	(0.001, 0.005)	0.97	0.75	6.92	(9.77)	928.78	(1253.12)
AFF	(η, α)	(0.001, 0.006)	0.97	0.73	6.70	(9.36)	952.18	(1267.24)
AFF	(η, α)	(0.001, 0.007)	0.97	0.71	6.50	(8.72)	832.73	(1163.07)
AFF	(η, α)	(0.001, 0.008)	0.97	0.71	6.57	(9.28)	793.05	(1095.04)
AFF	(η, α)	(0.001, 0.009)	0.97	0.70	6.49	(9.19)	782.56	(1051.46)
AFF	(η, α)	(0.001, 0.010)	0.97	0.68	6.55	(9.73)	729.64	(1004.34)
AFF	(η, α)	(0.001, 0.025)	0.98	0.59	6.27	(10.87)	416.94	(546.95)
AFF	(η, α)	(0.001, 0.050)	0.98	0.52	5.72	(10.40)	301.30	(379.49)

Table 46: Summary of detection efficiency for algorithms listed, over 750,000 observations with approximately 5000 changepoints, with $\delta = 2$ and B=50.

Algo	Params	Values	CCD	DNF	ARL1	(ARL1)	ARL0	(ARL0)
CUSUM	(k, h)	(0.25, 8.01)	0.98	0.73	6.97	(10.59)	281.57	(431.99)
CUSUM	(k, h)	(0.50, 4.77)	0.97	0.74	5.00	(7.77)	357.41	(582.96)
CUSUM	(k, h)	(0.75, 3.34)	0.95	0.75	4.82	(9.36)	440.62	(757.59)
CUSUM	(k, h)	(1.00, 2.52)	0.93	0.75	4.59	(9.48)	479.24	(848.11)
CUSUM	(k, h)	(1.25, 1.99)	0.91	0.75	4.60	(10.53)	526.88	(906.81)
CUSUM	(k, h)	(1.50, 1.61)	0.88	0.75	4.39	(10.19)	493.38	(855.65)
EWMA	(r, L)	(1.00, 3.090)	0.87	0.79	5.23	(10.83)	661.37	(987.72)
EWMA	(r, L)	(0.75, 3.087)	0.88	0.78	4.56	(10.64)	676.26	(1072.18)
EWMA	(r, L)	(0.50, 3.071)	0.91	0.78	4.52	(9.68)	638.20	(983.45)
EWMA	(r, L)	(0.40, 3.054)	0.93	0.78	4.90	(10.66)	564.67	(886.39)
EWMA	(r, L)	(0.30, 3.023)	0.95	0.77	4.57	(9.15)	506.76	(819.15)
EWMA	(r, L)	(0.25, 2.998)	0.95	0.77	4.53	(8.70)	502.91	(821.13)
EWMA	(r, L)	(0.20, 2.962)	0.96	0.77	4.47	(8.33)	493.38	(799.77)
EWMA	(r, L)	(0.10, 2.814)	0.97	0.56	5.18	(9.79)	407.02	(690.85)
EWMA	(r, L)	(0.05, 2.615)	1.00	0.34	1.23	(2.35)	356.06	(568.99)
EWMA	(r, L)	(0.03, 2.437)	1.00	0.34	1.11	(1.90)	312.45	(456.83)
AFF	(η, α)	(0.100, 0.005)	0.98	0.78	5.48	(8.76)	670.03	(972.68)
AFF	(η, α)	(0.100, 0.006)	0.98	0.76	5.53	(9.07)	576.34	(893.73)
AFF	(η, α)	(0.100, 0.007)	0.98	0.74	5.55	(9.07)	494.74	(765.15)
AFF	(η, α)	(0.100, 0.008)	0.98	0.73	5.55	(9.01)	396.80	(629.73)
AFF	(η, α)	(0.100, 0.009)	0.98	0.72	5.68	(9.53)	388.93	(634.65)
AFF	(η, α)	(0.100, 0.010)	0.98	0.70	5.78	(9.81)	352.77	(557.18)
AFF	(η, α)	(0.100, 0.025)	0.98	0.59	6.17	(11.53)	149.56	(199.33)
AFF	(η, α)	(0.100, 0.050)	0.99	0.52	5.77	(10.99)	87.67	(95.61)
AFF	(η, α)	(0.010, 0.005)	0.97	0.73	6.00	(10.57)	766.12	(1051.05)
AFF	(η, α)	(0.010, 0.006)	0.97	0.72	6.10	(11.02)	697.17	(1011.78)
AFF	(η, α)	(0.010, 0.007)	0.97	0.70	6.02	(10.89)	596.72	(895.37)
AFF	(η, α)	(0.010, 0.008)	0.97	0.68	6.13	(11.58)	594.55	(893.43)
AFF	(η, α)	(0.010, 0.009)	0.97	0.67	6.04	(11.42)	502.06	(720.98)
AFF	(η, α)	(0.010, 0.010)	0.97	0.66	6.06	(11.74)	463.64	(698.64)
AFF	(η, α)	(0.010, 0.025)	0.97	0.56	6.23	(12.20)	225.22	(308.82)
AFF	(η, α)	(0.010, 0.050)	0.98	0.49	5.96	(12.39)	146.05	(145.92)
AFF	(η, α)	(0.001, 0.005)	0.97	0.74	4.99	(8.50)	928.78	(1253.12)
AFF	(η, α)	(0.001, 0.006)	0.97	0.72	4.91	(8.40)	952.18	(1267.24)
AFF	(η, α)	(0.001, 0.007)	0.97	0.71	4.84	(8.25)	832.73	(1163.07)
AFF	(η, α)	(0.001, 0.008)	0.97	0.70	4.83	(8.35)	793.05	(1095.04)
AFF	(η, α)	(0.001, 0.009)	0.97	0.69	4.90	(8.93)	782.56	(1051.46)
AFF	(η, α)	(0.001, 0.010)	0.97	0.68	4.77	(8.60)	729.64	(1004.34)
AFF	(η, α)	(0.001, 0.025)	0.98	0.58	4.90	(10.19)	416.94	(546.95)
AFF	(η, α)	(0.001, 0.050)	0.98	0.51	4.45	(9.64)	301.30	(379.49)

Table 47: Summary of detection efficiency for algorithms listed, over 750,000 observations with approximately 5000 changepoints, with $\delta = 3$ and B=50.

Algo	Params	Values	CCD	DNF	ARL1	(ARL1)	ARL0	(ARL0)
CUSUM	(k, h)	(0.25, 8.01)	0.98	0.73	6.22	(10.06)	281.57	(431.99)
CUSUM	(k, h)	(0.50, 4.77)	0.96	0.74	4.41	(7.54)	357.41	(582.96)
CUSUM	(k, h)	(0.75, 3.34)	0.95	0.74	3.95	(7.84)	440.62	(757.59)
CUSUM	(k, h)	(1.00, 2.52)	0.93	0.75	3.57	(7.70)	479.24	(848.11)
CUSUM	(k, h)	(1.25, 1.99)	0.90	0.75	3.58	(8.71)	526.88	(906.81)
CUSUM	(k, h)	(1.50, 1.61)	0.87	0.75	3.12	(6.88)	493.38	(855.65)
EWMA	(r, L)	(1.00, 3.090)	0.86	0.79	3.78	(9.78)	661.37	(987.72)
EWMA	(r, L)	(0.75, 3.087)	0.87	0.79	3.43	(9.17)	676.26	(1072.18)
EWMA	(r, L)	(0.50, 3.071)	0.90	0.78	3.61	(8.95)	638.20	(983.45)
EWMA	(r, L)	(0.40, 3.054)	0.92	0.78	3.73	(8.70)	564.67	(886.39)
EWMA	(r, L)	(0.30, 3.023)	0.94	0.77	3.81	(8.15)	506.76	(819.15)
EWMA	(r, L)	(0.25, 2.998)	0.95	0.77	3.65	(7.24)	502.91	(821.13)
EWMA	(r, L)	(0.20, 2.962)	0.96	0.76	3.60	(6.82)	493.38	(799.77)
EWMA	(r, L)	(0.10, 2.814)	0.97	0.49	4.61	(9.96)	407.02	(690.85)
EWMA	(r, L)	(0.05, 2.615)	1.00	0.34	1.20	(2.07)	356.06	(568.99)
EWMA	(r, L)	(0.03, 2.437)	1.00	0.34	1.06	(1.31)	312.45	(456.83)
AFF	(η, α)	(0.100, 0.005)	0.98	0.77	5.30	(10.28)	670.03	(972.68)
AFF	(η, α)	(0.100, 0.006)	0.98	0.75	5.34	(10.42)	576.34	(893.73)
AFF	(η, α)	(0.100, 0.007)	0.98	0.74	5.40	(10.43)	494.74	(765.15)
AFF	(η, α)	(0.100, 0.008)	0.98	0.72	5.50	(10.84)	396.80	(629.73)
AFF	(η, α)	(0.100, 0.009)	0.98	0.71	5.71	(11.31)	388.93	(634.65)
AFF	(η, α)	(0.100, 0.010)	0.98	0.70	5.80	(11.53)	352.77	(557.18)
AFF	(η, α)	(0.100, 0.025)	0.98	0.59	5.96	(11.72)	149.56	(199.33)
AFF	(η, α)	(0.100, 0.050)	0.98	0.51	5.58	(11.20)	87.67	(95.61)
AFF	(η, α)	(0.010, 0.005)	0.97	0.73	5.09	(10.13)	766.12	(1051.05)
AFF	(η, α)	(0.010, 0.006)	0.97	0.71	5.24	(10.86)	697.17	(1011.78)
AFF	(η, α)	(0.010, 0.007)	0.96	0.70	5.40	(11.60)	596.72	(895.37)
AFF	(η, α)	(0.010, 0.008)	0.96	0.68	5.38	(11.56)	594.55	(893.43)
AFF	(η, α)	(0.010, 0.009)	0.96	0.67	5.29	(11.17)	502.06	(720.98)
AFF	(η, α)	(0.010, 0.010)	0.96	0.66	5.40	(11.77)	463.64	(698.64)
AFF	(η, α)	(0.010, 0.025)	0.96	0.55	6.01	(13.25)	225.22	(308.82)
AFF	(η, α)	(0.010, 0.050)	0.98	0.48	5.59	(12.49)	146.05	(145.92)
AFF	(η, α)	(0.001, 0.005)	0.97	0.74	4.11	(7.79)	928.78	(1253.12)
AFF	(η, α)	(0.001, 0.006)	0.97	0.72	4.15	(8.08)	952.18	(1267.24)
AFF	(η, α)	(0.001, 0.007)	0.97	0.71	4.12	(8.29)	832.73	(1163.07)
AFF	(η, α)	(0.001, 0.008)	0.97	0.69	4.09	(8.12)	793.05	(1095.04)
AFF	(η, α)	(0.001, 0.009)	0.97	0.68	4.16	(8.74)	782.56	(1051.46)
AFF	(η, α)	(0.001, 0.010)	0.97	0.67	4.19	(8.68)	729.64	(1004.34)
AFF	(η, α)	(0.001, 0.025)	0.97	0.57	4.34	(9.99)	416.94	(546.95)
AFF	(η, α)	(0.001, 0.050)	0.98	0.50	4.30	(10.39)	301.30	(379.49)

Table 48: Summary of detection efficiency for algorithms listed, over 750,000 observations with approximately 5000 changepoints, with $\delta = 4$ and B=50.

3.2.1 Discussion

These tables provide the same information as in Section 3.1, however now the tables are grouped by change size δ, rather than by algorithm and control parameters.

3.3 More tables for AFF as η varies

In this section, Table 49 provides further evidence, as shown in Table 1 in the main paper, that the value of the step-size η makes little difference to the performance of the AFF chaneg detector. For each choice of α, the CCD, DNF and ARL1 values are very similar for $\eta \in \{0.1, 0.01, 0.001\}$. However, the ARL0 values are slightly better for $\eta = 0.001$.

Algo	Params	Values	δ	CCD	DNF	ARL1	(ARL1)	ARL0	(ARL0)
AFF	(η, α)	(0.100, 0.005)	(0.25)	0.55	0.88	56.67	(39.80)	670.03	(972.68)
AFF	(η, α)	(0.010, 0.005)	(0.25)	0.58	0.86	56.65	(40.34)	766.12	(1051.05)
AFF	(η, α)	(0.001, 0.005)	(0.25)	0.59	0.86	55.30	(39.01)	928.78	(1253.12)
AFF	(η, α)	(0.100, 0.005)	(0.50)	0.86	0.84	40.61	(32.27)	670.03	(972.68)
AFF	(η, α)	(0.010, 0.005)	(0.50)	0.88	0.83	42.02	(32.92)	766.12	(1051.05)
AFF	(η, α)	(0.001, 0.005)	(0.50)	0.89	0.82	37.61	(30.12)	928.78	(1253.12)
AFF	(η, α)	(0.100, 0.005)	(0.75)	0.96	0.81	24.13	(20.46)	670.03	(972.68)
AFF	(η, α)	(0.010, 0.005)	(0.75)	0.96	0.79	25.03	(21.12)	766.12	(1051.05)
AFF	(η, α)	(0.001, 0.005)	(0.75)	0.95	0.79	21.86	(17.94)	928.78	(1253.12)
AFF	(η, α)	(0.010, 0.005)	(1.00)	0.97	0.77	16.47	(14.63)	766.12	(1051.05)
AFF	(η, α)	(0.100, 0.005)	(1.00)	0.97	0.80	15.24	(13.01)	670.03	(972.68)
AFF	(η, α)	(0.001, 0.005)	(1.00)	0.96	0.76	14.24	(12.08)	928.78	(1253.12)
AFF	(η, α)	(0.100, 0.005)	(1.50)	0.97	0.78	9.07	(9.76)	670.03	(972.68)
AFF	(η, α)	(0.010, 0.005)	(1.50)	0.97	0.74	10.00	(11.35)	766.12	(1051.05)
AFF	(η, α)	(0.001, 0.005)	(1.50)	0.97	0.75	8.91	(10.39)	928.78	(1253.12)
AFF	(η, α)	(0.100, 0.005)	(2.00)	0.98	0.78	7.06	(9.48)	670.03	(972.68)
AFF	(η, α)	(0.010, 0.005)	(2.00)	0.97	0.74	7.58	(9.99)	766.12	(1051.05)
AFF	(η, α)	(0.001, 0.005)	(2.00)	0.97	0.75	6.92	(9.77)	928.78	(1253.12)
AFF	(η, α)	(0.100, 0.005)	(3.00)	0.98	0.78	5.48	(8.76)	670.03	(972.68)
AFF	(η, α)	(0.010, 0.005)	(3.00)	0.97	0.73	6.00	(10.57)	766.12	(1051.05)
AFF	(η, α)	(0.001, 0.005)	(3.00)	0.97	0.74	4.99	(8.50)	928.78	(1253.12)
AFF	(η, α)	(0.100, 0.005)	(4.00)	0.98	0.77	5.30	(10.28)	670.03	(972.68)
AFF	(η, α)	(0.010, 0.005)	(4.00)	0.97	0.73	5.09	(10.13)	766.12	(1051.05)
AFF	(η, α)	(0.001, 0.005)	(4.00)	0.97	0.74	4.11	(7.79)	928.78	(1253.12)

Table 49: This table shows the results for AFF with $\alpha = 0.005$ and $\eta \in \{0.1, 0.01, 0.001\}$, for streams with approximately 5000 changepoints and different δ values, using burn-in length B=50. One notes that performance is similar for all values of η.

AFF	(η, α)	(0.100, 0.005)	0.85	0.82	27.23	(32.27)	670.04	(1018.23)
AFF	(η, α)	(0.010, 0.005)	0.86	0.79	27.12	(32.05)	819.36	(1162.97)
AFF	(η, α)	(0.001, 0.005)	0.86	0.78	24.89	(29.65)	987.68	(1336.78)
AFF	(η, α)	(0.100, 0.006)	0.86	0.80	26.87	(31.98)	564.99	(842.05)
AFF	(η, α)	(0.010, 0.006)	0.86	0.77	27.13	(32.23)	737.61	(1079.25)
AFF	(η, α)	(0.001, 0.006)	0.87	0.77	24.70	(29.58)	894.51	(1244.34)
AFF	(η, α)	(0.100, 0.007)	0.86	0.78	26.46	(31.77)	497.60	(745.82)
AFF	(η, α)	(0.010, 0.007)	0.87	0.76	26.31	(31.54)	659.02	(998.91)
AFF	(η, α)	(0.001, 0.007)	0.88	0.75	24.46	(29.40)	822.47	(1171.14)
AFF	(η, α)	(0.100, 0.008)	0.87	0.77	25.75	(31.00)	452.32	(706.86)
AFF	(η, α)	(0.010, 0.008)	0.87	0.73	25.78	(30.97)	577.88	(866.34)
AFF	(η, α)	(0.001, 0.008)	0.88	0.75	24.15	(29.15)	781.49	(1133.52)
AFF	(η, α)	(0.100, 0.009)	0.87	0.76	25.40	(30.80)	395.05	(604.51)
AFF	(η, α)	(0.010, 0.009)	0.87	0.72	25.18	(30.26)	528.15	(794.80)
AFF	(η, α)	(0.001, 0.009)	0.88	0.73	23.65	(28.54)	731.43	(1073.94)
AFF	(η, α)	(0.100, 0.010)	0.88	0.74	25.10	(30.55)	356.84	(548.64)
AFF	(η, α)	(0.010, 0.010)	0.88	0.71	24.96	(30.10)	495.66	(760.56)
AFF	(η, α)	(0.001, 0.010)	0.89	0.72	23.76	(29.14)	703.34	(1039.41)
AFF	(η, α)	(0.100, 0.025)	0.92	0.63	21.63	(27.52)	150.27	(194.32)
AFF	(η, α)	(0.010, 0.025)	0.92	0.61	21.19	(26.80)	234.74	(281.47)
AFF	(η, α)	(0.001, 0.025)	0.93	0.63	20.50	(25.80)	428.10	(638.33)
AFF	(η, α)	(0.100, 0.050)	0.96	0.55	17.80	(23.82)	87.08	(93.56)
AFF	(η, α)	(0.010, 0.050)	0.95	0.53	17.04	(23.02)	146.29	(151.69)
AFF	(η, α)	(0.001, 0.050)	0.96	0.55	17.26	(22.68)	279.61	(351.08)

Table 50: This table shows the results for AFF for different α, with $\eta \in \{0.1, 0.01, 0.001\}$, for streams with approximately 5000 changepoints with $\delta \in \{0.25, 0.5, 1, 3\}$, using burn-in length B=50. One notes that performance is similar for all values of η.

3.4 The effect of the burn-in length on SDRL0

In this section, Tables 51-53 show the effect of the size of the burn-in length on SDRL0 for CUSUM, EWMA and AFF. $B = \infty$ refers to the case when stream parameters are *known* (not estimated). Although results are only shown for one set of parameter values for each algorithm, a similar pattern is seen when other parameter pairs are used. The tables show that burn-in length around 300 yields similar SDRL0 value for when the parameter values are known and not estimated ($B = \infty$). This is in line with results from Jensen et al (2006).

Algo	Params	Values	Burn-in length	ARL0	(SDRL0)
CUSUM	(k, h)	(0.25, 8.01)	30	275.45	552.42
CUSUM	(k, h)	(0.25, 8.01)	50	307.25	608.12
CUSUM	(k, h)	(0.25, 8.01)	100	278.74	318.54
CUSUM	(k, h)	(0.25, 8.01)	200	319.96	369.09
CUSUM	(k, h)	(0.25, 8.01)	300	329.48	348.24
CUSUM	(k, h)	(0.25, 8.01)	400	321.59	333.60
CUSUM	(k, h)	(0.25, 8.01)	500	334.92	354.84
CUSUM	(k, h)	(0.25, 8.01)	∞	367.04	357.66

Table 51: ARL0 and SDRL0 values for CUSUM, estimated over 1000 trials.

Algo	Params	Values	Burn-in length	ARL0	(SDRL0)
EWMA	(r, L)	(0.25, 2.998)	30	626.98	1409.12
EWMA	(r, L)	(0.25, 2.998)	50	552.59	1043.89
EWMA	(r, L)	(0.25, 2.998)	100	463.38	611.54
EWMA	(r, L)	(0.25, 2.998)	200	486.51	596.22
EWMA	(r, L)	(0.25, 2.998)	300	498.06	574.30
EWMA	(r, L)	(0.25, 2.998)	400	503.21	565.25
EWMA	(r, L)	(0.25, 2.998)	500	492.09	526.76
EWMA	(r, L)	(0.25, 2.998)	∞	489.24	494.45

Table 52: ARL0 and SDRL0 values for EWMA, estimated over 1000 trials.

Algo	Params	Values	Burn-in length	ARL0	(SDRL0)
AFF	(η, α)	(0.01, 0.01)	30	388.27	643.66
AFF	(η, α)	(0.01, 0.01)	50	507.07	861.01
AFF	(η, α)	(0.01, 0.01)	100	521.01	674.08
AFF	(η, α)	(0.01, 0.01)	200	562.77	634.93
AFF	(η, α)	(0.01, 0.01)	300	606.78	689.55
AFF	(η, α)	(0.01, 0.01)	400	619.53	669.89
AFF	(η, α)	(0.01, 0.01)	500	627.70	665.65
AFF	(η, α)	(0.01, 0.01)	∞	665.00	712.42

Table 53: ARL0 and SDRL0 values for AFF, estimated over 1000 trials.

4 Fixed and adaptive forgetting factors

Section 4 in the main paper introduces the forgetting and adaptive factor formulations

$$\bar{x}_{N,\lambda} = \frac{1}{w_{N,\lambda}} \sum_{i=1}^{N} \lambda^{N-i} x_i, \qquad w_{N,\lambda} = \sum_{i=1}^{N} \lambda^{N-i}.$$

for a data stream $x_1, x_2, \ldots$. It is worth spending a bit of time to see how the value of λ affects the estimation of the forgetting factor mean $\bar{x}_{N,\lambda}$. Figure 5 shows, for a particular stream $x_1, x_2, \ldots x_{300}$ with a changepoint at $\tau = 100$ and $X_1, X_2, \ldots X_{100} \sim \mathrm{N}(0,1)$, $X_{101}, \ldots X_{300} \sim N(3,1)$, how $\bar{x}_{N,\lambda}$ behaves after choosing different values of λ. One notices that when $\lambda = 0.9$, the forgetting factor mean $\bar{x}_{N,0.9}$ reacts to the change quickly

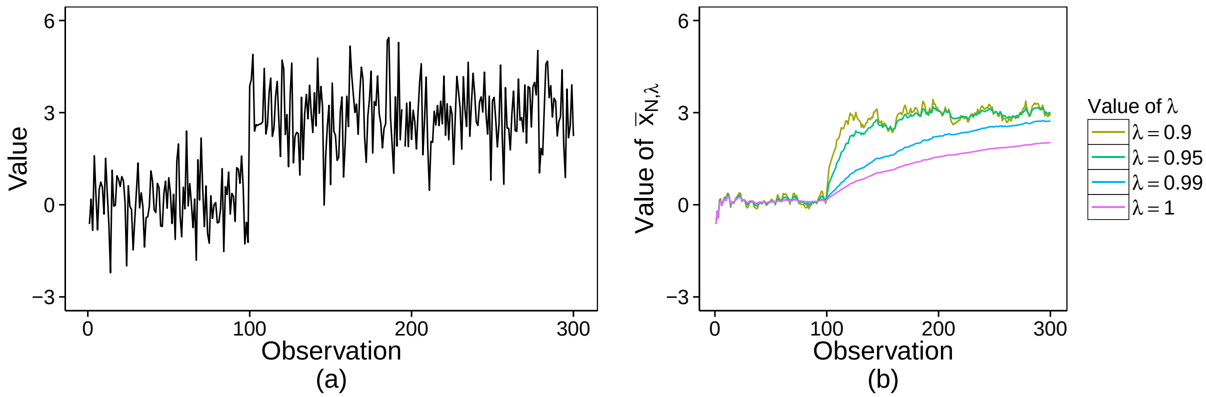

Figure 5: (a) A stream $x_1, x_2, \ldots, x_{300}$ sampled from $X_1, ..., X_{100} \sim \mathrm{N}(0,1)$, $X_{101}, \ldots, X_{300} \sim \mathrm{N}(3,1)$, and (b) the value of the fixed forgetting factor mean $\bar{x}_{N,\lambda}$ (on this stream) for different values of λ.

and is close to $\mu_2 = 3$ soon after the changepoint at $\tau = 100$. On the other hand, when $\lambda = 0.99$, the forgetting factor mean $\bar{x}_{N,0.99}$ reacts very slowly to the change, and is still not quite equal to $\mu_2 = 3$, even after observations x_{300}.

The case when $\lambda = 1$ is the situation when there is no forgetting, and the forgetting factor mean is simply the usual arithmetic mean, i.e. $\bar{x}_{N,1} = \bar{x}_N$. In this case, the reaction to the change is far slower than for $\lambda = 0.99$.

Figure 5 only shows the values of $\bar{x}_{N,\lambda}$, ofr a single stream. Figure 6 shows the average behaviour over 100 such streams.

One therefore sees that the smaller λ is, the faster it reacts to a change, and one would naturally think that the smaller the value of λ, the better. However, as discussed in Section 1.6 of this Supplementary Material, using λ smaller than $\lambda = 0.6$ can be counterproductive, and can lead to a large variance in the estimate of $\bar{X}_{N,\lambda}$.

Our proposed solution to use an *adaptive* forgetting factor attempts to avoid the problem of needing to select a value for λ, by using a data-driven approach to set the value of λ after each observation. A nice feature of the proposed scheme is that the adaptive forgetting factor $\overrightarrow{\lambda}$ will be close to 1 while the stream is in-control, but will drop in value after a change in order to allow to quickly forget the past regime and react to the change. Figure 7 is the same as Figure 6, but now the adaptive forgetting factor mean $\bar{x}_{N,\overrightarrow{\lambda}}$ has been included. One sees that the adaptive scheme reacts to the change more quickly than the fixed forgetting factor schemes.

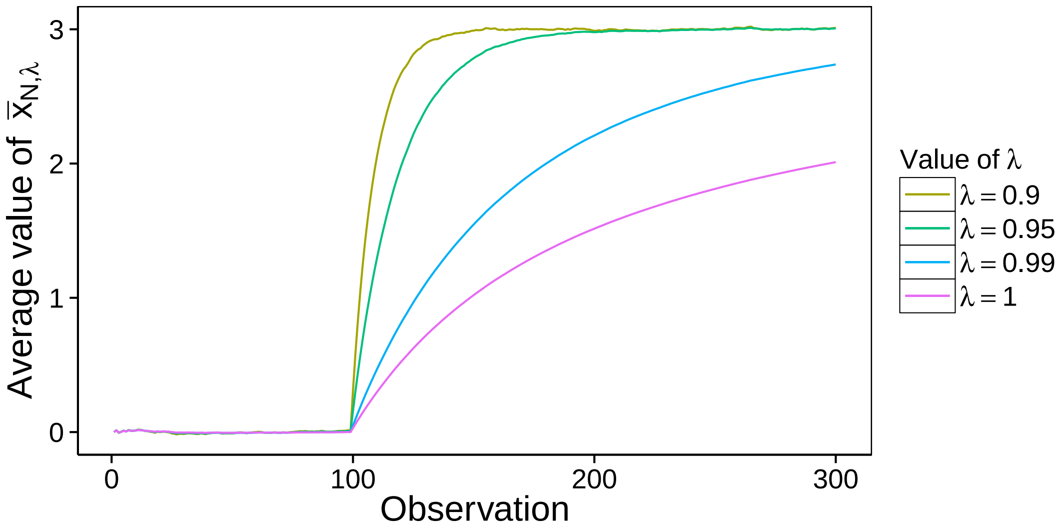

Figure 6: The average behaviour of $\bar{x}_{N,\lambda}$ for different values of λ, averaged over 100 streams $x_1, x_2, \ldots, x_{300}$ generated according to $X_1, ..., X_{100} \sim \mathrm{N}(0, 1)$, $X_{101}, \ldots, X_{300} \sim \mathrm{N}(3, 1)$.

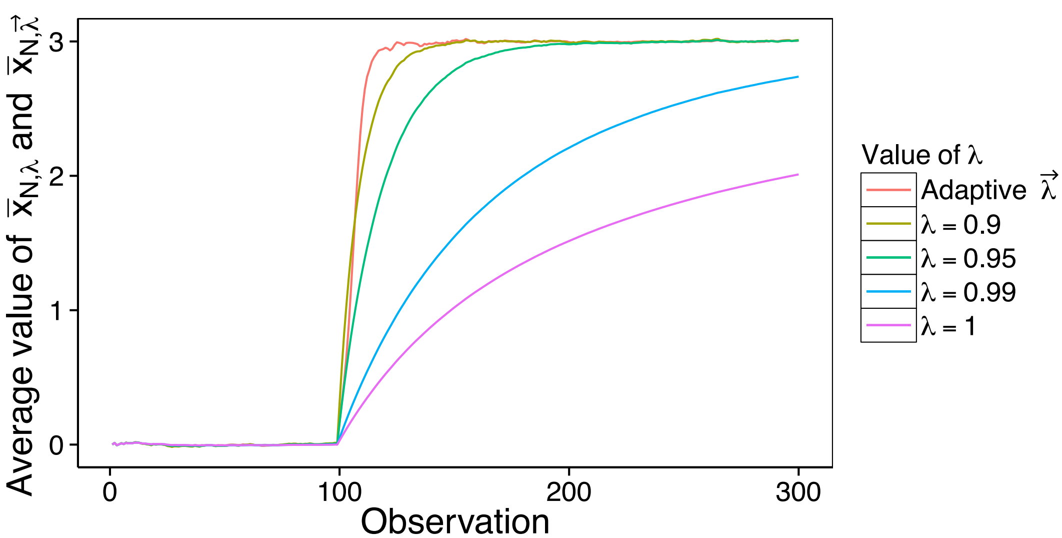

Figure 7: The average behaviour of $\bar{x}_{N,\lambda}$ for different values of λ, as well as $\bar{x}_{N,\vec{\lambda}}$, averaged over 100 streams $x_1, x_2, \ldots, x_{300}$ generated according to $X_1, ..., X_{100} \sim \mathrm{N}(0, 1)$, $X_{101}, \ldots, X_{300} \sim \mathrm{N}(3, 1)$.

5 Simulation study for both normally-distributed and gamma-distributed data and a comparison to Jiang-Shu-Apley

The following tables are an expanded simulation study, comparing CUSUM, EWMA, AFF and JSA (the Jiang-Shu-Apley method (Jiang et al, 2008). Simulations are performed on both normally-distributed, and gamma-distributed data. Note that the ten sets of parameters used for JSA were those values recommended in (Jiang et al, 2008).

The gamma-distributed data is generated in the same way as the normally distributed data, and the construction of the simulation study is described in detail in Section 5 of the main paper. For the gamma-distributed data, starting with a stream with mean $\mu = 1$ and $\sigma = 1$, which translates to a $\Gamma(1, 1)$ distribution, where $k = 1$ is the shape parameter and $\theta = 1$ is the scale parameter. After each change, $\mu_{new} = \mu_{old} \pm \delta$, where δ is in some set, e.g. $\{0.25, 0.5, 1, 3\}$. The parameter $\sigma = 1$ does not change. As long as $\mu > 0$ and $\sigma > 0$, then we can compute the parameters k and θ by using:

$$k = \mu^2/\theta^2$$
$$\theta = \sigma^2/\mu$$

In the event that $\mu_{new} = \mu_{old} - \delta$, and consequently $\mu_{new} \leq 0$, we rather set $\mu_{new} = \mu_{old} + \delta$, which is guaranteed to be positive, since $\mu_{old} > 0$ and $\delta > 0$.

The tables showing the various methods are presented below, even though the tables for CUSUM, EWMA and AFF are duplicates of those above. After the tables a discussion is provided analysing the resuls. To make it clear which distribution is used in each table, each table's caption contains “Data: NORMAL”, if normally-distributed data has been used, or “Data: GAMMA” if gamma-distributed data has been used. This is the only section where streams with gamma-distributed observations have been considered; all the other tables in the other sections of this Supplementary Material, and the paper, consider normally-distributed data.

5.1 Normally-distributed data, $\delta \in \{0.25, 0.5, 1, 3\}$

Algo	Params	Values	CCD	DNF	ARL1	(SDRL1)	ARL0	(SDRL0)
CUSUM	(k, h)	(0.25, 8.01)	0.90	0.77	24.19	(27.90)	285.25	(458.91)
CUSUM	(k, h)	(0.50, 4.77)	0.85	0.77	24.74	(30.54)	373.38	(612.43)
CUSUM	(k, h)	(0.75, 3.34)	0.80	0.78	26.09	(32.29)	454.60	(795.29)
CUSUM	(k, h)	(1.00, 2.52)	0.75	0.79	27.16	(34.17)	524.86	(905.54)
CUSUM	(k, h)	(1.25, 1.99)	0.70	0.80	29.96	(36.59)	538.56	(903.96)
CUSUM	(k, h)	(1.50, 1.61)	0.65	0.80	30.57	(36.85)	534.54	(890.49)

Table 54: Summary of detection efficiency for algorithms listed, over 750000 observations with 4998 detections, with $\delta \in \{0.25, 0.5, 1, 3\}$ with probabilities $\{0.25, 0.25, 0.25, 0.25\}$, with B=50. Data: NORMAL.

Algo	Params	Values	CCD	DNF	ARL1	(SDRL1)	ARL0	(SDRL0)
EWMA	(r, L)	(1.00, 3.090)	0.58	0.85	34.59	(38.96)	701.96	(1068.50)
EWMA	(r, L)	(0.75, 3.087)	0.64	0.83	32.14	(38.29)	654.44	(1000.28)
EWMA	(r, L)	(0.50, 3.071)	0.71	0.82	28.92	(35.42)	615.59	(969.71)
EWMA	(r, L)	(0.40, 3.054)	0.75	0.82	27.94	(34.78)	602.86	(948.69)
EWMA	(r, L)	(0.30, 3.023)	0.79	0.81	26.19	(32.88)	559.71	(892.78)
EWMA	(r, L)	(0.25, 2.998)	0.81	0.81	26.24	(33.22)	553.58	(884.68)
EWMA	(r, L)	(0.20, 2.962)	0.83	0.80	25.37	(32.20)	506.41	(847.37)
EWMA	(r, L)	(0.10, 2.814)	0.87	0.73	23.23	(29.68)	423.17	(704.08)
EWMA	(r, L)	(0.05, 2.615)	0.99	0.35	2.93	(10.41)	327.14	(561.52)

Table 55: Summary of detection efficiency for algorithms listed, over 750000 observations with 4998 detections, with $\delta \in \{0.25, 0.5, 1, 3\}$ with probabilities $\{0.25, 0.25, 0.25, 0.25\}$, with B=50. Data: NORMAL.

Algo	Params	Values	CCD	DNF	ARL1	(SDRL1)	ARL0	(SDRL0)
AFF	(η, α)	(0.100, 0.005)	0.85	0.82	27.26	(32.30)	670.04	(1018.23)
AFF	(η, α)	(0.100, 0.006)	0.86	0.80	26.88	(31.97)	564.99	(842.05)
AFF	(η, α)	(0.100, 0.007)	0.86	0.78	26.47	(31.76)	497.60	(745.82)
AFF	(η, α)	(0.100, 0.008)	0.87	0.77	25.79	(31.03)	452.32	(706.86)
AFF	(η, α)	(0.100, 0.009)	0.87	0.76	25.41	(30.79)	395.05	(604.51)
AFF	(η, α)	(0.100, 0.010)	0.88	0.74	25.06	(30.49)	356.84	(548.64)
AFF	(η, α)	(0.100, 0.025)	0.92	0.63	21.63	(27.52)	150.27	(194.32)
AFF	(η, α)	(0.100, 0.050)	0.96	0.55	17.83	(23.84)	87.08	(93.56)

Table 56: Summary of detection efficiency for algorithms listed, over 750000 observations with 4998 detections, with $\delta \in \{0.25, 0.5, 1, 3\}$ with probabilities $\{0.25, 0.25, 0.25, 0.25\}$, with B=50. Data: NORMAL.

Algo	Params	Values	CCD	DNF	ARL1	(SDRL1)	ARL0	(SDRL0)
AFF	(η, α)	(0.010, 0.005)	0.86	0.79	27.16	(32.12)	819.36	(1162.97)
AFF	(η, α)	(0.010, 0.006)	0.86	0.78	27.11	(32.19)	737.61	(1079.25)
AFF	(η, α)	(0.010, 0.007)	0.87	0.76	26.29	(31.53)	659.02	(998.91)
AFF	(η, α)	(0.010, 0.008)	0.87	0.73	25.71	(30.91)	577.88	(866.34)
AFF	(η, α)	(0.010, 0.009)	0.87	0.72	25.17	(30.24)	528.15	(794.80)
AFF	(η, α)	(0.010, 0.010)	0.88	0.71	24.98	(30.12)	495.66	(760.56)
AFF	(η, α)	(0.010, 0.025)	0.92	0.61	21.19	(26.83)	234.74	(281.47)
AFF	(η, α)	(0.010, 0.050)	0.95	0.53	17.05	(23.09)	146.29	(151.69)

Table 57: Summary of detection efficiency for algorithms listed, over 750000 observations with 4998 detections, with $\delta \in \{0.25, 0.5, 1, 3\}$ with probabilities $\{0.25, 0.25, 0.25, 0.25\}$, with B=50. Data: NORMAL.

Algo	Params	Values	CCD	DNF	ARL1	(SDRL1)	ARL0	(SDRL0)
AFF	(η, α)	(0.001, 0.005)	0.86	0.78	25.00	(29.82)	987.68	(1336.78)
AFF	(η, α)	(0.001, 0.006)	0.87	0.77	24.67	(29.52)	894.51	(1244.34)
AFF	(η, α)	(0.001, 0.007)	0.88	0.75	24.49	(29.42)	822.47	(1171.14)
AFF	(η, α)	(0.001, 0.008)	0.88	0.75	24.12	(29.12)	781.49	(1133.52)
AFF	(η, α)	(0.001, 0.009)	0.88	0.73	23.67	(28.60)	731.43	(1073.94)
AFF	(η, α)	(0.001, 0.010)	0.89	0.72	23.75	(29.12)	703.34	(1039.41)
AFF	(η, α)	(0.001, 0.025)	0.93	0.63	20.44	(25.74)	428.10	(638.33)
AFF	(η, α)	(0.001, 0.050)	0.96	0.55	17.34	(22.84)	279.61	(351.08)

Table 58: Summary of detection efficiency for algorithms listed, over 750000 observations with 4998 detections, with $\delta \in \{0.25, 0.5, 1, 3\}$ with probabilities $\{0.25, 0.25, 0.25, 0.25\}$, with B=50. Data: NORMAL.

Algo	Params	Values	CCD	DNF	ARL1	(SDRL1)	ARL0	(SDRL0)
JSA	($\widehat{\delta}_{\min}$, β, γ, ζ)	(0.500, 0.20, 1.50, 6.06)	0.12	0.94	27.57	(31.33)	489.99	(642.11)
JSA	($\widehat{\delta}_{\min}$, β, γ, ζ)	(0.500, 0.20, 2.00, 5.11)	0.13	0.90	27.31	(31.83)	285.47	(344.19)
JSA	($\widehat{\delta}_{\min}$, β, γ, ζ)	(0.500, 0.20, 2.50, 4.63)	0.13	0.87	28.09	(33.04)	222.09	(265.45)
JSA	($\widehat{\delta}_{\min}$, β, γ, ζ)	(0.500, 0.20, 3.00, 4.43)	0.13	0.87	28.78	(33.76)	203.55	(244.33)
JSA	($\widehat{\delta}_{\min}$, β, γ, ζ)	(0.500, 0.20, 4.00, 4.35)	0.13	0.86	27.93	(33.73)	198.73	(242.35)
JSA	($\widehat{\delta}_{\min}$, β, γ, ζ)	(1.000, 0.30, 1.50, 5.05)	0.12	0.88	26.29	(33.10)	236.44	(268.91)
JSA	($\widehat{\delta}_{\min}$, β, γ, ζ)	(1.000, 0.30, 2.00, 4.73)	0.12	0.85	27.23	(34.10)	190.84	(215.01)
JSA	($\widehat{\delta}_{\min}$, β, γ, ζ)	(1.000, 0.30, 2.50, 4.50)	0.13	0.84	26.41	(33.38)	160.89	(176.55)
JSA	($\widehat{\delta}_{\min}$, β, γ, ζ)	(1.000, 0.30, 3.00, 4.39)	0.13	0.83	26.38	(33.64)	152.14	(166.09)
JSA	($\widehat{\delta}_{\min}$, β, γ, ζ)	(1.000, 0.30, 4.00, 4.34)	0.13	0.82	27.24	(34.07)	150.25	(165.05)

Table 59: Summary of detection efficiency for algorithms listed, over 750000 observations with 4998 detections, with $\delta \in \{0.25, 0.5, 1, 3\}$ with probabilities $\{0.25, 0.25, 0.25, 0.25\}$, with B=50. Data: NORMAL.

5.1.1 Discussion

Tables 54-59 are more extensive versions of Table 2 in the main paper. The new information is contained in Table 59, which clearly shows that JSA is not suitable for the continuous monitoring scenario we present. The chief concern is that the CCD value is very low. This could be due to several factors (a) there being only a (relatively) short time between changepoints, and so there is not sufficient time for the method to detect certain changepoints, (b) the use of esitmated parameters may be affecting the algorithms effectiveness (it original paper assumes pre-change parameters are known), (c) the size of some changepoints are too small, e.g. $\delta = 0.25$, or $\delta = 0.5$.

In order to address point (c), since $\widehat{\delta}_{\min} \in \{0.5, 1.0\}$, we run the algorithms again using streams which are generated using changepoints in the set $\delta \in \{1, 2, 3, 4\}$, all of which are larger than (or equal to) $\widehat{\delta}_{\min} \in \{0.5, 1.0\}$. These simulations are detailed in Sections 5.3 and 5.4 below.

5.2 Gamma-distributed data, $\delta \in \{0.25, 0.5, 1, 3\}$

Algo	Params	Values	CCD	DNF	ARL1	(SDRL1)	ARL0	(SDRL0)
CUSUM	(k, h)	(0.25, 8.01)	0.90	0.77	24.39	(27.97)	355.99	(711.13)
CUSUM	(k, h)	(0.50, 4.77)	0.86	0.78	24.50	(30.86)	315.29	(718.33)
CUSUM	(k, h)	(0.75, 3.34)	0.81	0.79	25.82	(32.99)	166.17	(408.27)
CUSUM	(k, h)	(1.00, 2.52)	0.75	0.79	27.89	(35.04)	110.45	(270.61)
CUSUM	(k, h)	(1.25, 1.99)	0.71	0.79	29.27	(35.80)	85.01	(177.92)
CUSUM	(k, h)	(1.50, 1.61)	0.67	0.79	30.23	(36.81)	79.18	(168.94)

Table 60: Summary of detection efficiency for algorithms listed, over 750000 observations with 5002 detections, with $\delta \in$ {0.25, 0.5, 1, 3} with probabilities {0.25, 0.25, 0.25, 0.25}, with B=50. Data: GAMMA.

Algo	Params	Values	CCD	DNF	ARL1	(SDRL1)	ARL0	(SDRL0)
EWMA	(r, L)	(1.00, 3.090)	0.59	0.84	34.27	(39.34)	82.35	(170.56)
EWMA	(r, L)	(0.75, 3.087)	0.65	0.83	31.10	(37.24)	94.69	(208.90)
EWMA	(r, L)	(0.50, 3.071)	0.73	0.82	28.31	(35.16)	136.03	(311.39)
EWMA	(r, L)	(0.40, 3.054)	0.76	0.82	26.97	(34.02)	176.04	(412.21)
EWMA	(r, L)	(0.30, 3.023)	0.79	0.82	25.85	(32.94)	263.02	(623.97)
EWMA	(r, L)	(0.25, 2.998)	0.82	0.82	25.42	(32.45)	321.92	(714.31)
EWMA	(r, L)	(0.20, 2.962)	0.83	0.81	24.97	(31.86)	417.29	(888.51)
EWMA	(r, L)	(0.10, 2.814)	0.90	0.58	18.66	(27.35)	661.10	(1229.39)
EWMA	(r, L)	(0.05, 2.615)	1.00	0.35	2.18	(8.32)	499.96	(904.12)

Table 61: Summary of detection efficiency for algorithms listed, over 750000 observations with 5002 detections, with $\delta \in$ {0.25, 0.5, 1, 3} with probabilities {0.25, 0.25, 0.25, 0.25}, with B=50. Data: GAMMA.

Algo	Params	Values	CCD	DNF	ARL1	(SDRL1)	ARL0	(SDRL0)
AFF	(η, α)	(0.100, 0.005)	0.85	0.82	25.89	(30.78)	629.63	(1100.11)
AFF	(η, α)	(0.100, 0.006)	0.86	0.81	25.27	(30.25)	582.82	(1039.06)
AFF	(η, α)	(0.100, 0.007)	0.87	0.79	25.30	(30.32)	533.22	(979.02)
AFF	(η, α)	(0.100, 0.008)	0.88	0.78	25.08	(30.12)	491.99	(912.66)
AFF	(η, α)	(0.100, 0.009)	0.88	0.76	24.88	(29.90)	459.28	(861.58)
AFF	(η, α)	(0.100, 0.010)	0.89	0.75	24.39	(29.44)	425.23	(790.21)
AFF	(η, α)	(0.100, 0.025)	0.93	0.63	21.19	(26.53)	230.75	(439.56)
AFF	(η, α)	(0.100, 0.050)	0.96	0.55	18.00	(23.70)	132.35	(246.26)

Table 62: Summary of detection efficiency for algorithms listed, over 750000 observations with 5002 detections, with $\delta \in$ {0.25, 0.5, 1, 3} with probabilities {0.25, 0.25, 0.25, 0.25}, with B=50. Data: GAMMA.

Algo	Params	Values	CCD	DNF	ARL1	(SDRL1)	ARL0	(SDRL0)
AFF	(η, α)	(0.010, 0.005)	0.86	0.80	26.98	(31.72)	740.16	(1088.95)
AFF	(η, α)	(0.010, 0.006)	0.87	0.78	26.13	(30.98)	680.05	(1034.86)
AFF	(η, α)	(0.010, 0.007)	0.87	0.76	25.58	(30.61)	623.33	(956.83)
AFF	(η, α)	(0.010, 0.008)	0.88	0.74	25.27	(30.20)	573.37	(905.30)
AFF	(η, α)	(0.010, 0.009)	0.89	0.73	25.01	(30.37)	534.41	(838.78)
AFF	(η, α)	(0.010, 0.010)	0.89	0.72	24.74	(30.21)	507.05	(803.60)
AFF	(η, α)	(0.010, 0.025)	0.92	0.61	21.00	(26.41)	301.29	(534.51)
AFF	(η, α)	(0.010, 0.050)	0.95	0.53	17.22	(22.74)	174.44	(247.31)

Table 63: Summary of detection efficiency for algorithms listed, over 750000 observations with 5002 detections, with $\delta \in$ {0.25, 0.5, 1, 3} with probabilities {0.25, 0.25, 0.25, 0.25}, with B=50. Data: GAMMA.

Algo	Params	Values	CCD	DNF	ARL1	(SDRL1)	ARL0	(SDRL0)
AFF	(η, α)	(0.001, 0.005)	0.87	0.79	24.81	(29.91)	1034.23	(1369.45)
AFF	(η, α)	(0.001, 0.006)	0.88	0.77	24.57	(29.65)	964.41	(1300.42)
AFF	(η, α)	(0.001, 0.007)	0.89	0.76	23.92	(29.04)	905.21	(1243.37)
AFF	(η, α)	(0.001, 0.008)	0.89	0.75	23.79	(28.92)	851.50	(1175.53)
AFF	(η, α)	(0.001, 0.009)	0.89	0.74	23.24	(28.21)	807.31	(1129.67)
AFF	(η, α)	(0.001, 0.010)	0.90	0.73	23.31	(28.47)	752.74	(1060.50)
AFF	(η, α)	(0.001, 0.025)	0.93	0.62	20.06	(24.93)	452.77	(683.34)
AFF	(η, α)	(0.001, 0.050)	0.96	0.55	17.21	(22.69)	291.52	(387.35)

Table 64: Summary of detection efficiency for algorithms listed, over 750000 observations with 5002 detections, with $\delta \in \{0.25, 0.5, 1, 3\}$ with probabilities $\{0.25, 0.25, 0.25, 0.25\}$, with B=50. Data: GAMMA.

Algo	Params	Values	CCD	DNF	ARL1	(SDRL1)	ARL0	(SDRL0)
JSA	($\widehat{\delta}_{\min}$, β, γ, ζ)	(0.500, 0.20, 1.50, 6.06)	0.09	0.91	31.95	(31.27)	72.70	(77.02)
JSA	($\widehat{\delta}_{\min}$, β, γ, ζ)	(0.500, 0.20, 2.00, 5.11)	0.09	0.87	30.73	(33.09)	55.33	(53.53)
JSA	($\widehat{\delta}_{\min}$, β, γ, ζ)	(0.500, 0.20, 2.50, 4.63)	0.09	0.84	31.04	(34.84)	50.74	(48.19)
JSA	($\widehat{\delta}_{\min}$, β, γ, ζ)	(0.500, 0.20, 3.00, 4.43)	0.10	0.83	28.58	(32.72)	52.28	(49.17)
JSA	($\widehat{\delta}_{\min}$, β, γ, ζ)	(0.500, 0.20, 4.00, 4.35)	0.10	0.82	27.42	(31.04)	59.54	(56.36)
JSA	($\widehat{\delta}_{\min}$, β, γ, ζ)	(1.000, 0.30, 1.50, 5.05)	0.09	0.86	26.75	(31.44)	56.60	(61.29)
JSA	($\widehat{\delta}_{\min}$, β, γ, ζ)	(1.000, 0.30, 2.00, 4.73)	0.09	0.83	27.35	(32.78)	52.20	(57.51)
JSA	($\widehat{\delta}_{\min}$, β, γ, ζ)	(1.000, 0.30, 2.50, 4.50)	0.10	0.81	26.99	(33.41)	50.41	(54.70)
JSA	($\widehat{\delta}_{\min}$, β, γ, ζ)	(1.000, 0.30, 3.00, 4.39)	0.10	0.81	27.35	(34.20)	51.03	(52.71)
JSA	($\widehat{\delta}_{\min}$, β, γ, ζ)	(1.000, 0.30, 4.00, 4.34)	0.10	0.80	27.44	(33.66)	56.34	(60.80)

Table 65: Summary of detection efficiency for algorithms listed, over 750000 observations with 5002 detections, with $\delta \in \{0.25, 0.5, 1, 3\}$ with probabilities $\{0.25, 0.25, 0.25, 0.25\}$, with B=50. Data: GAMMA.

5.2.1 Discussion

One can try and compare the methods by extracting the following cases into Table 66 below. Between CUSUM, EWMA and AFF, we pick parameter pairs that match the CCD, DNF and ARL1 as closely as possible. The ARL1 for AFF is slightly worse than CUSUM and EWMA for this case, but it is still very close. EWMA has slightly better DNF than AFF, but worse CCD. The CCD for AFF is at as good/better than the CCD for CUSUM and EWMA. However, although these three metrics (CCD, DNF and ARL1) are all very similar for CUSUM, EWMA and AFF, the ARL0 for AFF is far better than the ARL0 of CUSUM and EWMA. This suggests that the model misspecification of using gamma-distributed data, rather than normally-distributed data, does not affect AFF as much as it affects CUSUM and EWMA.

One set of values for JSA has been included, but looking at Table 65, it seems that JSA is very negatively affected by the normality assumption being violated (compare to Table 59, above, which uses normally-distributed data).

Algo	Params	Values	CCD	DNF	ARL1	(SDRL1)	ARL0	(SDRL0)
CUSUM	(k, h)	(0.50, 4.77)	0.86	0.78	24.50	(30.86)	315.29	(718.33)
CUSUM	(k, h)	(0.75, 3.34)	0.81	0.79	25.82	(32.99)	166.17	(408.27)
EWMA	(r, L)	(0.25, 2.998)	0.82	0.82	25.42	(32.45)	321.92	(714.31)
AFF	(η, α)	(0.010, 0.005)	0.86	0.80	26.98	(31.72)	740.16	(1088.95)
JSA	($\widehat{\delta}_{\min}$, β, γ, ζ)	(0.500, 0.20, 1.50, 6.06)	0.09	0.91	31.95	(31.27)	72.70	(77.02)

Table 66: Summary of detection efficiency for algorithms listed, over 750000 observations with 5002 detections, with $\delta \in \{0.25, 0.5, 1, 3\}$ with probabilities $\{0.25, 0.25, 0.25, 0.25\}$, with B=50. Data: GAMMA.

Another point to note is that the ARL0 is very negatively affected for values of $k \geq 0.75$; for example, CUSUM ($k = 0.75$, $h = 3.34$) has ARL0=166.17 for gamma-distributed data, but ARL0=454.60 for normally-dsitributed data. See Table 54 for more details.

5.3 Normally-distributed data, $\delta \in \{1, 2, 3, 4\}$

Algo	Params	Values	CCD	DNF	ARL1	(SDRL1)	ARL0	(SDRL0)
CUSUM	(k, h)	(0.25, 8.01)	0.98	0.74	9.12	(11.89)	285.25	(458.91)
CUSUM	(k, h)	(0.50, 4.77)	0.96	0.74	7.71	(11.45)	373.38	(612.43)
CUSUM	(k, h)	(0.75, 3.34)	0.94	0.75	7.89	(13.09)	454.60	(795.29)
CUSUM	(k, h)	(1.00, 2.52)	0.92	0.76	8.45	(14.66)	524.86	(905.54)
CUSUM	(k, h)	(1.25, 1.99)	0.90	0.76	9.71	(17.73)	538.56	(903.96)
CUSUM	(k, h)	(1.50, 1.61)	0.86	0.76	11.14	(20.71)	534.54	(890.49)

Table 67: Summary of detection efficiency for algorithms listed, over 750000 observations with 4998 detections, with $\delta \in$ {1, 2, 3, 4} with probabilities {0.25, 0.25, 0.25, 0.25}, with B=50. Data: NORMAL.

Algo	Params	Values	CCD	DNF	ARL1	(SDRL1)	ARL0	(SDRL0)
EWMA	(r, L)	(1.00, 3.090)	0.80	0.81	15.46	(26.70)	701.96	(1068.50)
EWMA	(r, L)	(0.75, 3.087)	0.87	0.80	12.40	(22.79)	654.44	(1000.28)
EWMA	(r, L)	(0.50, 3.071)	0.90	0.78	9.35	(16.71)	615.59	(969.71)
EWMA	(r, L)	(0.40, 3.054)	0.92	0.78	8.52	(15.01)	602.86	(948.69)
EWMA	(r, L)	(0.30, 3.023)	0.94	0.77	7.75	(13.20)	559.71	(892.78)
EWMA	(r, L)	(0.25, 2.998)	0.95	0.77	7.48	(12.14)	553.58	(884.68)
EWMA	(r, L)	(0.20, 2.962)	0.95	0.77	7.24	(11.59)	506.41	(847.37)
EWMA	(r, L)	(0.10, 2.814)	0.96	0.58	7.37	(11.93)	423.17	(704.08)
EWMA	(r, L)	(0.05, 2.615)	1.00	0.34	1.44	(3.50)	327.14	(561.52)

Table 68: Summary of detection efficiency for algorithms listed, over 750000 observations with 4998 detections, with $\delta \in$ {1, 2, 3, 4} with probabilities {0.25, 0.25, 0.25, 0.25}, with B=50. Data: NORMAL.

Algo	Params	Values	CCD	DNF	ARL1	(SDRL1)	ARL0	(SDRL0)
AFF	(η, α)	(0.100, 0.005)	0.98	0.78	8.35	(11.38)	670.04	(1018.23)
AFF	(η, α)	(0.100, 0.006)	0.98	0.76	8.21	(11.28)	564.99	(842.05)
AFF	(η, α)	(0.100, 0.007)	0.98	0.75	8.12	(11.37)	497.60	(745.82)
AFF	(η, α)	(0.100, 0.008)	0.97	0.73	8.18	(11.81)	452.32	(706.86)
AFF	(η, α)	(0.100, 0.009)	0.98	0.72	8.37	(12.31)	395.05	(604.51)
AFF	(η, α)	(0.100, 0.010)	0.97	0.71	8.32	(12.40)	356.84	(548.64)
AFF	(η, α)	(0.100, 0.025)	0.97	0.60	7.98	(12.87)	150.27	(194.32)
AFF	(η, α)	(0.100, 0.050)	0.98	0.52	7.21	(12.54)	87.08	(93.56)

Table 69: Summary of detection efficiency for algorithms listed, over 750000 observations with 4998 detections, with $\delta \in$ {1, 2, 3, 4} with probabilities {0.25, 0.25, 0.25, 0.25}, with B=50. Data: NORMAL.

Algo	Params	Values	CCD	DNF	ARL1	(SDRL1)	ARL0	(SDRL0)
AFF	(η, α)	(0.010, 0.005)	0.97	0.74	8.75	(12.55)	819.36	(1162.97)
AFF	(η, α)	(0.010, 0.006)	0.96	0.72	8.69	(12.85)	737.61	(1079.25)
AFF	(η, α)	(0.010, 0.007)	0.96	0.71	8.73	(13.34)	659.02	(998.91)
AFF	(η, α)	(0.010, 0.008)	0.96	0.69	8.68	(13.45)	577.88	(866.34)
AFF	(η, α)	(0.010, 0.009)	0.96	0.68	8.64	(13.35)	528.15	(794.80)
AFF	(η, α)	(0.010, 0.010)	0.97	0.67	8.50	(13.32)	495.66	(760.56)
AFF	(η, α)	(0.010, 0.025)	0.97	0.56	8.09	(13.15)	234.74	(281.47)
AFF	(η, α)	(0.010, 0.050)	0.98	0.49	7.36	(13.19)	146.29	(151.69)

Table 70: Summary of detection efficiency for algorithms listed, over 750000 observations with 4998 detections, with $\delta \in$ {1, 2, 3, 4} with probabilities {0.25, 0.25, 0.25, 0.25}, with B=50. Data: NORMAL.

Algo	Params	Values	CCD	DNF	ARL1	(SDRL1)	ARL0	(SDRL0)
AFF	(η, α)	(0.001, 0.005)	0.97	0.75	7.68	(11.60)	987.68	(1336.78)
AFF	(η, α)	(0.001, 0.006)	0.97	0.73	7.55	(11.23)	894.51	(1244.34)
AFF	(η, α)	(0.001, 0.007)	0.97	0.72	7.39	(10.99)	822.47	(1171.14)
AFF	(η, α)	(0.001, 0.008)	0.97	0.71	7.43	(11.25)	781.49	(1133.52)
AFF	(η, α)	(0.001, 0.009)	0.97	0.70	7.21	(10.68)	731.43	(1073.94)
AFF	(η, α)	(0.001, 0.010)	0.97	0.68	7.32	(11.44)	703.34	(1039.41)
AFF	(η, α)	(0.001, 0.025)	0.97	0.59	6.87	(11.70)	428.10	(638.33)
AFF	(η, α)	(0.001, 0.050)	0.98	0.51	6.21	(11.45)	279.61	(351.08)

Table 71: Summary of detection efficiency for algorithms listed, over 750000 observations with 4998 detections, with $\delta \in \{1, 2, 3, 4\}$ with probabilities $\{0.25, 0.25, 0.25, 0.25\}$, with B=50. Data: NORMAL.

Algo	Params	Values	CCD	DNF	ARL1	(SDRL1)	ARL0	(SDRL0)
JSA	$(\widehat{\delta}_{\min}, \beta, \gamma, \zeta)$	(0.500, 0.20, 1.50, 6.06)	0.09	0.93	12.34	(14.59)	489.99	(642.11)
JSA	$(\widehat{\delta}_{\min}, \beta, \gamma, \zeta)$	(0.500, 0.20, 2.00, 5.11)	0.09	0.90	13.44	(21.19)	285.47	(344.19)
JSA	$(\widehat{\delta}_{\min}, \beta, \gamma, \zeta)$	(0.500, 0.20, 2.50, 4.63)	0.09	0.88	13.74	(23.42)	222.09	(265.45)
JSA	$(\widehat{\delta}_{\min}, \beta, \gamma, \zeta)$	(0.500, 0.20, 3.00, 4.43)	0.10	0.86	14.24	(25.08)	203.55	(244.33)
JSA	$(\widehat{\delta}_{\min}, \beta, \gamma, \zeta)$	(0.500, 0.20, 4.00, 4.35)	0.10	0.86	13.09	(22.70)	198.73	(242.35)
JSA	$(\widehat{\delta}_{\min}, \beta, \gamma, \zeta)$	(1.000, 0.30, 1.50, 5.05)	0.09	0.87	11.60	(20.14)	236.44	(268.91)
JSA	$(\widehat{\delta}_{\min}, \beta, \gamma, \zeta)$	(1.000, 0.30, 2.00, 4.73)	0.10	0.84	13.45	(24.21)	190.84	(215.01)
JSA	$(\widehat{\delta}_{\min}, \beta, \gamma, \zeta)$	(1.000, 0.30, 2.50, 4.50)	0.09	0.82	13.23	(24.89)	160.89	(176.55)
JSA	$(\widehat{\delta}_{\min}, \beta, \gamma, \zeta)$	(1.000, 0.30, 3.00, 4.39)	0.10	0.81	13.25	(25.14)	152.14	(166.09)
JSA	$(\widehat{\delta}_{\min}, \beta, \gamma, \zeta)$	(1.000, 0.30, 4.00, 4.34)	0.10	0.80	13.51	(25.69)	150.25	(165.05)

Table 72: Summary of detection efficiency for algorithms listed, over 750000 observations with 4998 detections, with $\delta \in \{1, 2, 3, 4\}$ with probabilities $\{0.25, 0.25, 0.25, 0.25\}$, with B=50. Data: NORMAL.

5.3.1 Discussion

Again, we extract values from the tables above into Table 73 below. Again, the JSA method is behaving differently to the other methods, with low CCD, so we restrict out attention to CUSUM, EWMA and AFF. Two sets of CUSUM and two sets of EWMA values have been extracted, which are similar to the set for AFF. The two CUSUM sets closely match AFF for CCD, DNF and ARL1. The EWMA sets closely match the AFF set for CCD, although DNF experiences a huge drop between the two parameter choices, while the ARL1 is slightly better than that of AFF. However, AFF again has substantially better ARL0 than CUSUM and EWMA. Of course, the ARL0 values are the same as in Section 5.1 above, since the set of change sizes δ does not affect the ARL0 values (and both section use normally-distributed data).

Algo	Params	Values	CCD	DNF	ARL1	(SDRL1)	ARL0	(SDRL0)
CUSUM	(k, h)	(0.25, 8.01)	0.98	0.74	9.12	(11.89)	285.25	(458.91)
CUSUM	(k, h)	(0.50, 4.77)	0.96	0.74	7.71	(11.45)	373.38	(612.43)
EWMA	(r, L)	(0.20, 2.962)	0.95	0.77	7.24	(11.59)	506.41	(847.37)
EWMA	(r, L)	(0.10, 2.814)	0.96	0.58	7.37	(11.93)	423.17	(704.08)
AFF	(η, α)	(0.010, 0.005)	0.97	0.74	8.75	(12.55)	819.36	(1162.97)
JSA	$(\widehat{\delta}_{\min}, \beta, \gamma, \zeta)$	(0.500, 0.20, 1.50, 6.06)	0.09	0.93	12.34	(14.59)	489.99	(642.11)

Table 73: Summary of detection efficiency for algorithms listed, over 750000 observations with 4998 detections, with $\delta \in \{1, 2, 3, 4\}$ with probabilities $\{0.25, 0.25, 0.25, 0.25\}$, with B=50. Data: NORMAL.

5.4 Gamma-distributed data, $\delta \in \{1, 2, 3, 4\}$

Algo	Params	Values	CCD	DNF	ARL1	(SDRL1)	ARL0	(SDRL0)
CUSUM	(k, h)	(0.25, 8.01)	0.98	0.74	8.93	(11.93)	355.99	(711.13)
CUSUM	(k, h)	(0.50, 4.77)	0.96	0.74	7.66	(11.49)	315.29	(718.33)
CUSUM	(k, h)	(0.75, 3.34)	0.94	0.74	7.74	(12.67)	166.17	(408.27)
CUSUM	(k, h)	(1.00, 2.52)	0.92	0.75	8.48	(14.77)	110.45	(270.61)
CUSUM	(k, h)	(1.25, 1.99)	0.89	0.75	9.99	(18.23)	85.01	(177.92)
CUSUM	(k, h)	(1.50, 1.61)	0.86	0.76	11.01	(20.38)	79.18	(168.94)

Table 74: Summary of detection efficiency for algorithms listed, over 750000 observations with 4997 detections, with $\delta \in$ {1, 2, 3, 4} with probabilities {0.25, 0.25, 0.25, 0.25}, with B=50. Data: GAMMA.

Algo	Params	Values	CCD	DNF	ARL1	(SDRL1)	ARL0	(SDRL0)
EWMA	(r, L)	(1.00, 3.090)	0.79	0.80	15.69	(27.35)	82.35	(170.56)
EWMA	(r, L)	(0.75, 3.087)	0.86	0.80	11.90	(21.49)	94.69	(208.90)
EWMA	(r, L)	(0.50, 3.071)	0.90	0.78	9.38	(16.82)	136.03	(311.39)
EWMA	(r, L)	(0.40, 3.054)	0.92	0.78	8.58	(15.08)	176.04	(412.21)
EWMA	(r, L)	(0.30, 3.023)	0.94	0.77	7.77	(13.08)	263.02	(623.97)
EWMA	(r, L)	(0.25, 2.998)	0.95	0.77	7.40	(12.03)	321.92	(714.31)
EWMA	(r, L)	(0.20, 2.962)	0.96	0.76	7.33	(11.88)	417.29	(888.51)
EWMA	(r, L)	(0.10, 2.814)	0.98	0.67	7.04	(10.11)	661.10	(1229.39)
EWMA	(r, L)	(0.05, 2.615)	1.00	0.35	1.91	(5.45)	499.96	(904.12)

Table 75: Summary of detection efficiency for algorithms listed, over 750000 observations with 4997 detections, with $\delta \in$ {1, 2, 3, 4} with probabilities {0.25, 0.25, 0.25, 0.25}, with B=50. Data: GAMMA.

Algo	Params	Values	CCD	DNF	ARL1	(SDRL1)	ARL0	(SDRL0)
AFF	(η, α)	(0.100, 0.005)	0.98	0.78	8.37	(11.58)	629.63	(1100.11)
AFF	(η, α)	(0.100, 0.006)	0.98	0.76	8.45	(12.12)	582.82	(1039.06)
AFF	(η, α)	(0.100, 0.007)	0.98	0.75	8.37	(12.15)	533.22	(979.02)
AFF	(η, α)	(0.100, 0.008)	0.98	0.73	8.28	(12.13)	491.99	(912.66)
AFF	(η, α)	(0.100, 0.009)	0.98	0.72	8.35	(12.34)	459.28	(861.58)
AFF	(η, α)	(0.100, 0.010)	0.98	0.71	8.40	(12.59)	425.23	(790.21)
AFF	(η, α)	(0.100, 0.025)	0.98	0.60	8.05	(13.03)	230.75	(439.56)
AFF	(η, α)	(0.100, 0.050)	0.98	0.52	7.52	(12.71)	132.35	(246.26)

Table 76: Summary of detection efficiency for algorithms listed, over 750000 observations with 4997 detections, with $\delta \in$ {1, 2, 3, 4} with probabilities {0.25, 0.25, 0.25, 0.25}, with B=50. Data: GAMMA.

Algo	Params	Values	CCD	DNF	ARL1	(SDRL1)	ARL0	(SDRL0)
AFF	(η, α)	(0.010, 0.005)	0.97	0.74	8.74	(12.29)	740.16	(1088.95)
AFF	(η, α)	(0.010, 0.006)	0.97	0.72	8.70	(12.39)	680.05	(1034.86)
AFF	(η, α)	(0.010, 0.007)	0.97	0.71	8.44	(12.06)	623.33	(956.83)
AFF	(η, α)	(0.010, 0.008)	0.97	0.69	8.66	(12.95)	573.37	(905.30)
AFF	(η, α)	(0.010, 0.009)	0.97	0.68	8.69	(13.21)	534.41	(838.78)
AFF	(η, α)	(0.010, 0.010)	0.96	0.67	8.67	(13.37)	507.05	(803.60)
AFF	(η, α)	(0.010, 0.025)	0.97	0.56	8.55	(14.40)	301.29	(534.51)
AFF	(η, α)	(0.010, 0.050)	0.98	0.49	7.60	(13.28)	174.44	(247.31)

Table 77: Summary of detection efficiency for algorithms listed, over 750000 observations with 4997 detections, with $\delta \in$ {1, 2, 3, 4} with probabilities {0.25, 0.25, 0.25, 0.25}, with B=50. Data: GAMMA.

Algo	Params	Values	CCD	DNF	ARL1	(SDRL1)	ARL0	(SDRL0)
AFF	(η, α)	(0.001, 0.005)	0.97	0.74	7.63	(10.66)	1034.23	(1369.45)
AFF	(η, α)	(0.001, 0.006)	0.97	0.73	7.59	(10.87)	964.41	(1300.42)
AFF	(η, α)	(0.001, 0.007)	0.97	0.71	7.56	(11.01)	905.21	(1243.37)
AFF	(η, α)	(0.001, 0.008)	0.98	0.70	7.64	(11.48)	851.50	(1175.53)
AFF	(η, α)	(0.001, 0.009)	0.98	0.69	7.46	(11.26)	807.31	(1129.67)
AFF	(η, α)	(0.001, 0.010)	0.97	0.67	7.32	(11.14)	752.74	(1060.50)
AFF	(η, α)	(0.001, 0.025)	0.97	0.59	7.28	(12.73)	452.77	(683.34)
AFF	(η, α)	(0.001, 0.050)	0.98	0.51	6.66	(12.79)	291.52	(387.35)

Table 78: Summary of detection efficiency for algorithms listed, over 750000 observations with 4997 detections, with $\delta \in \{1, 2, 3, 4\}$ with probabilities $\{0.25, 0.25, 0.25, 0.25\}$, with B=50. Data: GAMMA.

Algo	Params	Values	CCD	DNF	ARL1	(SDRL1)	ARL0	(SDRL0)
JSA	$(\widehat{\delta}_{\min}, \beta, \gamma, \zeta)$	(0.500, 0.20, 1.50, 6.06)	0.14	0.94	14.27	(19.16)	72.70	(77.02)
JSA	$(\widehat{\delta}_{\min}, \beta, \gamma, \zeta)$	(0.500, 0.20, 2.00, 5.11)	0.15	0.91	13.45	(20.38)	55.33	(53.53)
JSA	$(\widehat{\delta}_{\min}, \beta, \gamma, \zeta)$	(0.500, 0.20, 2.50, 4.63)	0.15	0.88	14.07	(22.25)	50.74	(48.19)
JSA	$(\widehat{\delta}_{\min}, \beta, \gamma, \zeta)$	(0.500, 0.20, 3.00, 4.43)	0.16	0.87	13.85	(22.20)	52.28	(49.17)
JSA	$(\widehat{\delta}_{\min}, \beta, \gamma, \zeta)$	(0.500, 0.20, 4.00, 4.35)	0.16	0.86	13.47	(21.58)	59.54	(56.36)
JSA	$(\widehat{\delta}_{\min}, \beta, \gamma, \zeta)$	(1.000, 0.30, 1.50, 5.05)	0.15	0.87	12.67	(21.72)	56.60	(61.29)
JSA	$(\widehat{\delta}_{\min}, \beta, \gamma, \zeta)$	(1.000, 0.30, 2.00, 4.73)	0.16	0.85	12.64	(21.72)	52.20	(57.51)
JSA	$(\widehat{\delta}_{\min}, \beta, \gamma, \zeta)$	(1.000, 0.30, 2.50, 4.50)	0.16	0.84	13.25	(23.61)	50.41	(54.70)
JSA	$(\widehat{\delta}_{\min}, \beta, \gamma, \zeta)$	(1.000, 0.30, 3.00, 4.39)	0.16	0.83	14.14	(25.47)	51.03	(52.71)
JSA	$(\widehat{\delta}_{\min}, \beta, \gamma, \zeta)$	(1.000, 0.30, 4.00, 4.34)	0.16	0.82	13.47	(24.74)	56.34	(60.80)

Table 79: Summary of detection efficiency for algorithms listed, over 750000 observations with 4997 detections, with $\delta \in \{1, 2, 3, 4\}$ with probabilities $\{0.25, 0.25, 0.25, 0.25\}$, with B=50. Data: GAMMA.

5.4.1 Discussion

The results are similar for gamma-distributed data with $\delta \in \{1, 2, 3, 4\}$ in this section, when compared to those in Section 5.2 with $\delta \in \{0.25, 0.50, 1, 3\}$: while we can compare CUSUM, EWMA and AFF by matching the metrics DNF, CCF and ARL1, AFF has higher ARL0 than either CUSUM or EWMA. It is interesting to note that the CCD values for AFF seem fairly consistent, even though α is increasing, and the DNF decreases. As before, the JSA method underperforms, with very low CCD (although it has slightly increased compared to the previous sections).

6 A review of the Jiang-Shu-Apley method

This section contains a brief review of the Jiang-Shu-Apley method, which is fully described in Jiang et al (2008).

Suppose one wishes to monitor a stream $x_1, x_2, \dots$, which is initially generated from $X_0, X_1, \dots \sim \mathrm{N}(0,1)$ before experiencing a change at time τ. One starts by defining $\hat{\delta}_t$, e_t and the Huber function $\phi_{\beta,\gamma}$:

$$\hat{\delta}_0 = 0 \tag{21}$$

$$e_t = x_t - \hat{\delta}_{t-1}, \text{ for } t = 1, 2, \dots \tag{22}$$

$$\hat{\delta}_t = \hat{\delta}_{t-1} + \phi_{\beta,\gamma}(e_t) \tag{23}$$

where

$$\phi_{\beta,\gamma}(e) = \begin{cases} e + (1-\beta)\gamma, & \text{if } e < -\gamma, \\ \beta\gamma, & \text{if } |e| \leq \gamma, \\ e - (1-\beta)\gamma, & \text{if } e > \gamma. \end{cases} \tag{24}$$

and where β and γ are control parameters. Next, for another control parameter $\hat{\delta}^+_{\min} > 0$, we define $\hat{\delta}^+_t$:

$$\hat{\delta}^+_t = \max\{\hat{\delta}^+_{\min}, \hat{\delta}_t\}. \tag{25}$$

So, from a stream $x_1, x_2, \dots$, we get a stream $\hat{\delta}^+_1, \hat{\delta}^+_2, \dots$. We next define the function w to simply be:

$$w(x) = x \tag{26}$$

i.e. the identity function. It seems that other choices could be used for w, but Jiang et al (2008) favours using the identity function here. Now, the CUSUMs Z^+_t and Z^-_t can be defined:

$$Z^+_t = \max\{0, Z^+_{t-1} + w(\hat{\delta}^+_t)(x_t - \hat{\delta}^+_t/2)\} \tag{27}$$

$$Z^-_t = \max\{0, Z^-_{t-1} + w(\hat{\delta}^-_t)(x_t - \hat{\delta}^-_t/2)\} \tag{28}$$

It is not explicitly specified in the paper, but $\hat{\delta}^-_{\min} = -\hat{\delta}^+_{\min}$, and $\hat{\delta}^-_t = \min\{\hat{\delta}^-_{\min}, \hat{\delta}_t\}$.

Now, a change is signalled at time t' if $Z^+_{t'} > \zeta$ or $Z^-_{t'} < -\zeta$, for some control parameter $\zeta > 0$.

Recall that this formulation assumes that the stream is $\mathrm{N}(0,1)$-distributed. If the stream $\mathrm{N}(\mu,1)$ distributed, with $\mu \neq 0$, then a slight adjustment is needed, and all the above should be replace x_t with $y_t = x_t - \mu$.

7 Running CUSUM and EWMA on the foreign exchange data

Following a suggestion from an anonymous reviewer, we run CUSUM and EWMA on the CHF/GBP foreign exchange data used in Section 6 of the main paper, and compare them with PELT visually for the first 10,000 observations in the figures below. It appears that the different parameter choices can have a strong effect on the number of changepoints detected (although this is to be expected).

We also see how many of the 373 changepoints detected by PELT on the full dataset of $\sim 330,000$ observations are detected by each method, and compare that to AFF as the benchmark.

Note that AFF with $\alpha = 0.005$ detects 795 changepoints, of which 154 are within 10 observations of a PELT-detected changepoints. That means that, on this real dataset, we get for AFF:

$$\text{CCD} = 154/373 \approx 0.413$$
$$\text{DNF} = 154/795 \approx 0.194$$

where CCD is the proportion of (true) changepoints correctly detected, and DNF is the proportion of (detected) changepoints that are not false detections. These metrics are described in more detail in the main paper in Section 2.2.2.

7.1 CUSUM

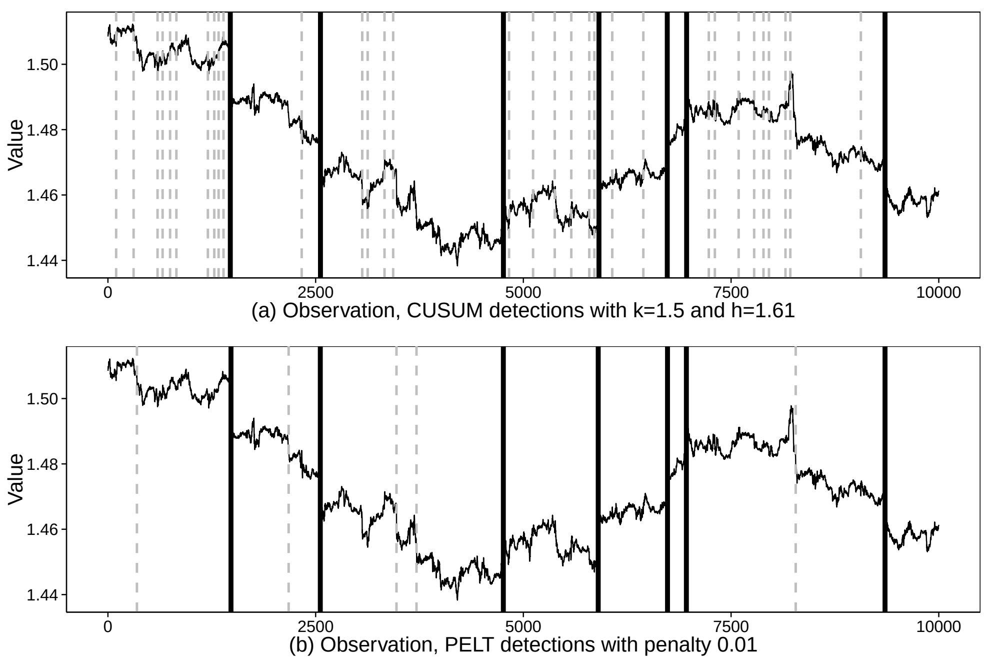

Figure 8: Change detection on a CHF/GBP data stream using (a) CUSUM (above) and (b) PELT (below). The raw data stream is plotted with the detected changepoints indicated by the vertical lines, with solid black lines indicating that both schemes detect that changepoint (within 10 observations of each other), and grey dashed lines indicating that the changepoint is not detected by the other method.

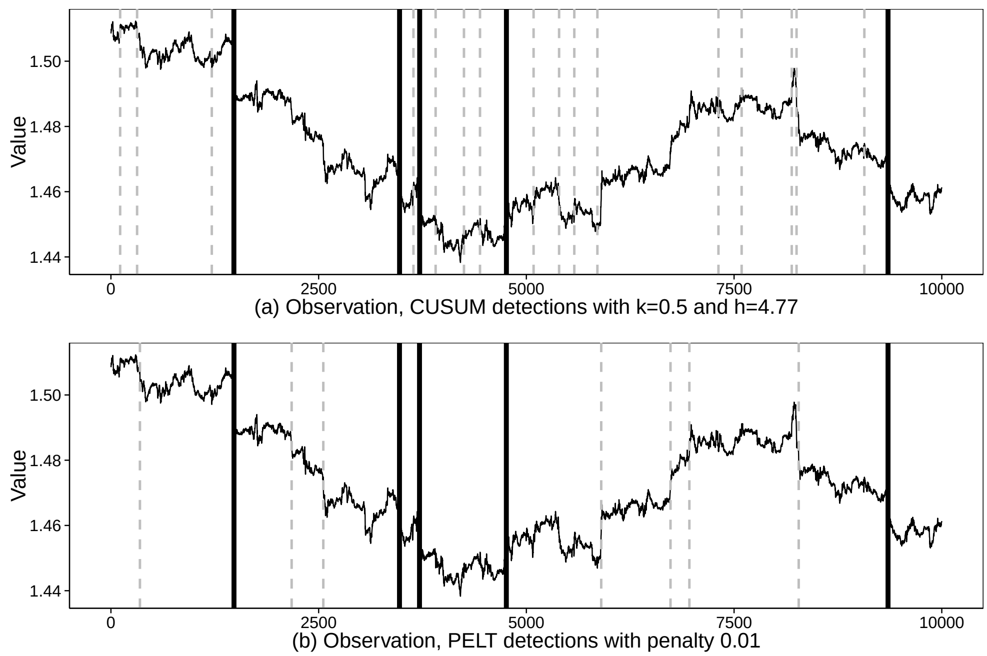

Figure 9: Change detection on a CHF/GBP data stream using (a) CUSUM (above) and (b) PELT (below). The raw data stream is plotted with the detected changepoints indicated by the vertical lines, with solid black lines indicating that both schemes detect that changepoint (within 10 observations of each other), and grey dashed lines indicating that the changepoint is not detected by the other method.

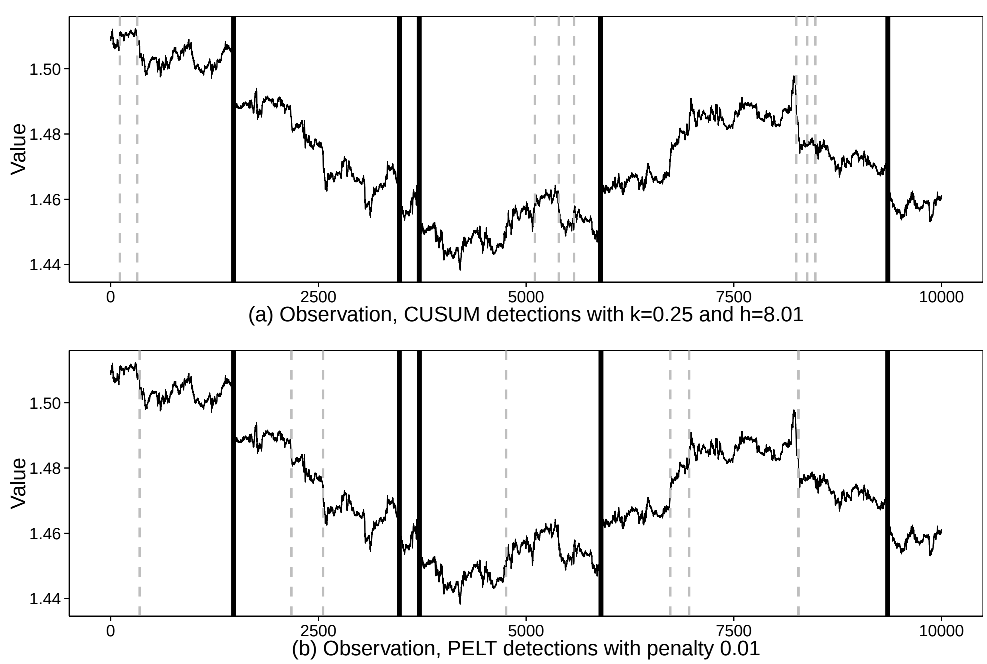

Figure 10: Change detection on a CHF/GBP data stream using (a) CUSUM (above) and (b) PELT (below). The raw data stream is plotted with the detected changepoints indicated by the vertical lines, with solid black lines indicating that both schemes detect that changepoint (within 10 observations of each other), and grey dashed lines indicating that the changepoint is not detected by the other method.

7.2 EWMA

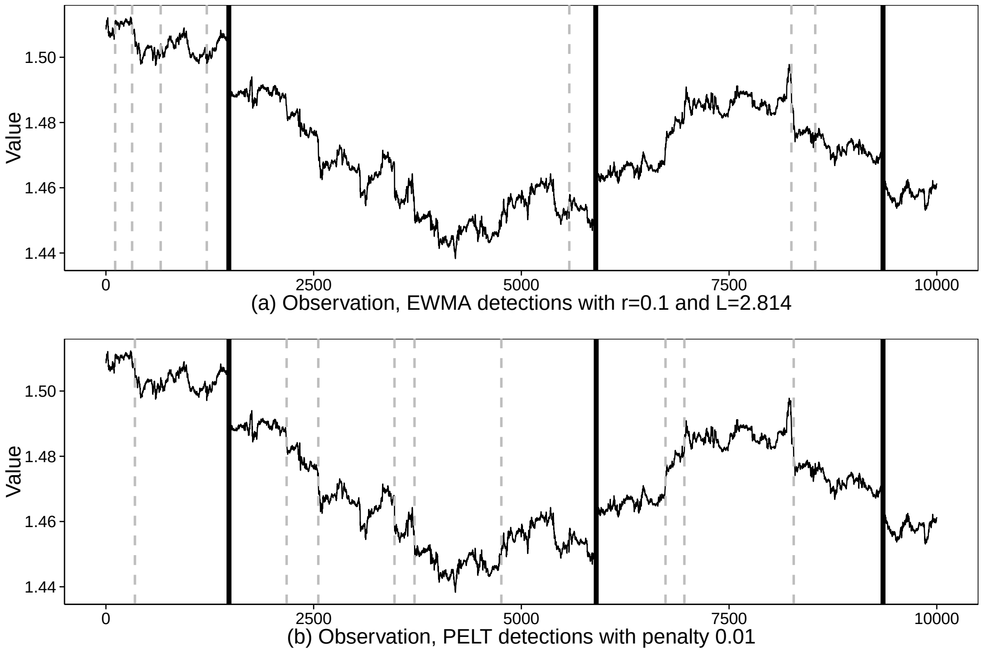

Figure 11: Change detection on a CHF/GBP data stream using (a) EWMA (above) and (b) PELT (below). The raw data stream is plotted with the detected changepoints indicated by the vertical lines, with solid black lines indicating that both schemes detect that changepoint (within 10 observations of each other), and grey dashed lines indicating that the changepoint is not detected by the other method.

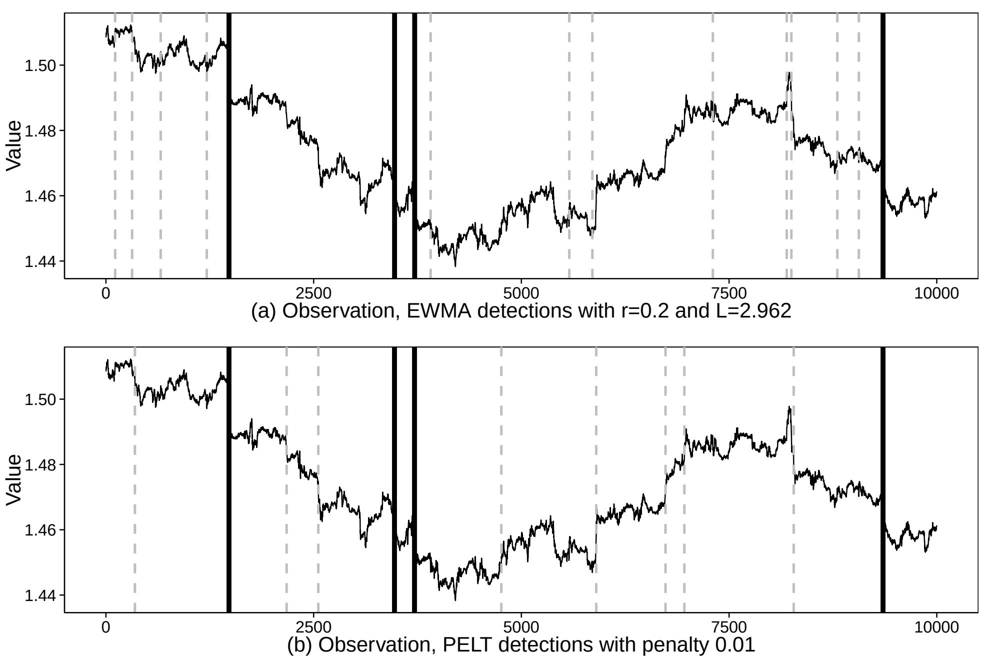

Figure 12: Change detection on a CHF/GBP data stream using (a) EWMA (above) and (b) PELT (below). The raw data stream is plotted with the detected changepoints indicated by the vertical lines, with solid black lines indicating that both schemes detect that changepoint (within 10 observations of each other), and grey dashed lines indicating that the changepoint is not detected by the other method.

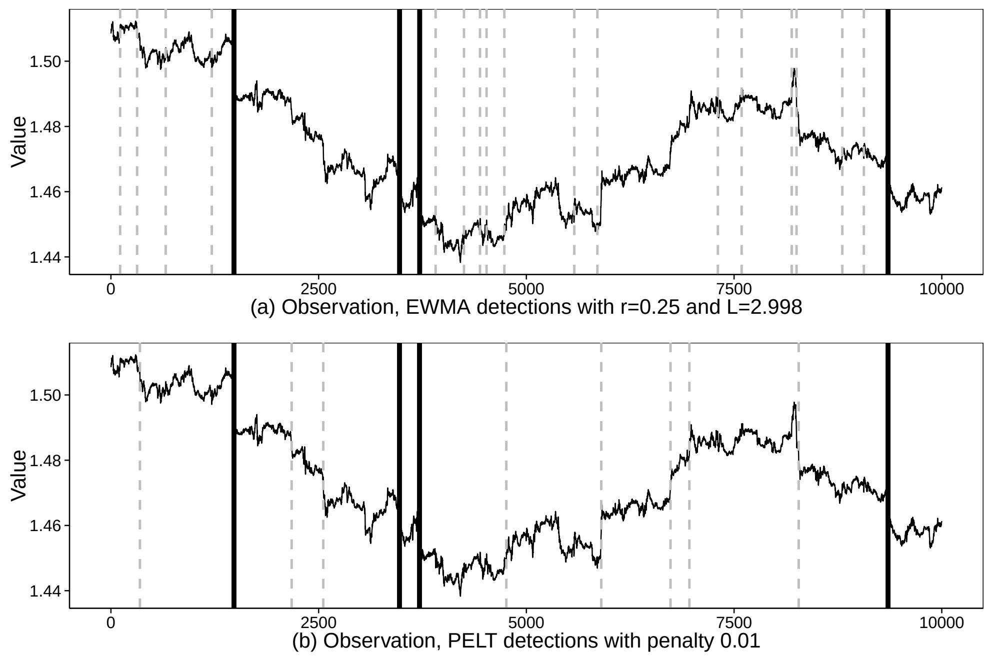

Figure 13: Change detection on a CHF/GBP data stream using (a) EWMA (above) and (b) PELT (below). The raw data stream is plotted with the detected changepoints indicated by the vertical lines, with solid black lines indicating that both schemes detect that changepoint (within 10 observations of each other), and grey dashed lines indicating that the changepoint is not detected by the other method.

7.3 Evaluation on full dataset

Table 80 evaluates how CUSUM and EWMA compare to AFF, when evaluated on the full CHF/GBP dataset of over 330,000 observations. In each case, the CCD is computed as the proportion of the "true" changepoints (detected by PELT) that are also detected by CUSUM/EWMA/AFF, while DNF is computed as the proportion of changepoints detected by CUSUM/EWMA/AFF that are "true" changepoints (detected by PELT).

We see that while AFF with $\alpha = 0.005$ has the highest CCD value, and clearly outperforms CUSUM-(1.51, 1.61) (for both CCD and DNF), EWMA has slightly higher DNF values in general. It is difficult to draw conclusions about which algorithm is performing better. However, as the main thrust of the argument in the paper discusses, we cannot choose control parameters after the fact; we would have needed to choose either CUSUM or EWMA, and a particular parameter pair, before monitoring the data. Although EWMA looks relatively stable, performance for CUSUM varies substantially. On the other hand, setting the single control parameter for AFF is relatively simple.

Algo	Parameters	Values	CCD	DNF
CUSUM	(k, h)	(1.51, 1.61)	0.386	0.149
CUSUM	(k, h)	(0.50, 4.77)	0.300	0.230
CUSUM	(k, h)	(0.25, 8.01)	0.164	0.285
EWMA	(r, L)	(0.10, 2.814)	0.321	0.254
EWMA	(r, L)	(0.20, 2.962)	0.354	0.211
EWMA	(r, L)	(0.25, 2.998)	0.343	0.225
AFF	α	0.005	0.413	0.194

Table 80: The CCD and DNF for CUSUM and EWMA, for a selection of parameter values, compared to AFF. The CCD is computed as the proportion of the PELT-detected changepoints that are detected by the CUSUM/EWMA/AFF, while DNF is computed as the proportion of changepoints detected by CUSUM/EWMA/AFF that are true changepoints (detected by PELT).

8 A note on the creation of $\overrightarrow{\lambda}$ in the R figures

Note that it does not currently seem to be possible in R to create the LaTeX symbol λ with an $\rightarrow$ above it (for use with axis labels), i.e. the symbol $\overrightarrow{\lambda}$, which is used extensively throughtout this paper as the adaptive forgetting factor. In order to get around this restriction, I noted that it is possible to create $\bar{\lambda}$ in R. I then used Inkscape to add a "$>$" ending to the "$-$" to turn $\bar{\lambda}$ into $\overrightarrow{\lambda}$. While this is not the cleanest solution, it unfortunately seemed to be the only one available to me at this time (June 2016).

9 Acknowledgements

Earlier versions of the derivations and figures in this supplementary material originally appeared in the unpublished PhD thesis, Bodenham (2014).

References

Bodenham DA (2014) Adaptive estimation with change detection for streaming data. PhD thesis, Imperial College London

Jensen WA, Jones-Farmer L, Champ CW, Woodall WH, et al (2006) Effects of parameter estimation on control chart properties: a literature review. Journal of Quality Technology 38(4):349–364

Jiang W, Shu W, Apley DW (2008) Adaptive cusum procedures with ewma-based shift estimators. IIE Transactions 40(10):992–1003